A Bayesian Approach to Dose-Finding Studies for Cancer Therapies: Producing Personalised Procedures While Incorporating Information from Later Cycles of Therapy

Thesis for the degree of PhD
July 2014

Karen Sinclair MSc, BSc (hons)

Department of Mathematics and Statistics

Lancaster University

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgments

I would like to thank the MRC who funded this project.

I would like to thank Professor Anne Whitehead for her supervision and constant support throughout this project. Her advice and contribution was invaluable, as was her encouragement and reassurance.

I would also like to thank the rest of the Medical and Pharmaceutical Statistics (MPS) Research Unit for their discussions and advice, and Lancaster University's Mathematics and Statistics Department for allowing me to continue my training and education for 7 years.

Dr Xavier Paoletti provided me with data from a published analysis of 38 Phase I trials [1] and also provided some insight into the data.

Discussions with clinicians from The Christie hospital in Manchester have provided insight into desired improvements for Phase I dose-finding studies from a clinical perspective. This added extra motivation into the development for personalised procedures. In particular Dr. Alistair Greystoke provided data and insight from a Phase I study [17], Professor Malcolm Ranson allowed me to attend two clinic days for which enrolled patients in a Phase I study were attending, and allowed me to study the protocol, and Professor Andrew Renehan organised the collaboration with the Christie Hospital.

I would like to express my appreciation to my fellow PhD colleagues and my admiration of them. Without them the past 3 years would not have seemed so easy and enjoyable. The reassurance each and every one of them gave could only come from those with exactly the same experience and understanding and I am eternally grateful.

Finally I would like to directly thank Andy, Dad, Clare and Richard. You have all supported me unreservedly. I love you and am forever thankful.

Declaration

The work in this thesis is my own work which has been guided by the supervision of Professor Anne Whitehead.

Work from this thesis has been published in Statistics in Medicine [4] (early view online version).

Abstract

Bayesian model based approaches for Phase I dose-finding studies are popular procedures to implement due to the efficiency of updating information sequentially after accruing information. Traditionally dose-finding studies for cancer treatments focus on the occurrence of a patient's first dose limiting toxicity in the first cycle of therapy.

This thesis develops a Bayesian decision procedure featuring an Interval-Censored Survival model to incorporate information from multiple cycles of therapy. The use of data from multiple cycles of therapy should produce more precise estimates of target doses to recommend for further investigation in later phases of drug development, in a shorter amount of time.

An increasingly desired approach in dose-finding procedures is to provide personalised procedures to target therapy to individual tolerances. Features such as allowing intra-patient dose adjustments and incorporating baseline characteristics to investigate the underlying drug tolerance of population subgroups are investigated within the use of the Interval-Censored Survival Decision Procedure (ICSDP).

The inclusion of time-varying covariates is also possible when using the ICSDP, which is investigated through including lower grade toxicities as a marker for tolerance. Individual target doses can be estimated, but the analysis of dose limiting toxicities alone provides a population target dose to recommend for further investigation.

Results show the ICSDP as an efficient approach to use when observing a patient's first dose limiting toxicity. Target doses are estimated with good precision, comparable to or better than existing designs for dose-finding, and are generally obtained in a shorter amount of time. Multiple target doses can be produced for different subgroups of the population when baseline characteristics are used and intra-patient dose adjustments are possible between cycles. When using intra-patient adjustments based on observation of lower grade toxicities, personalised dose-escalations lead to estimates of individual target doses and a population target dose with good precision in a reduced amount of time.

Contents

1. Introduction 1
2. Literature Review of Existing Phase I Dose-Finding Procedures 7
   2.1 Traditional Rule Based Designs 7
      2.1.1 3+3 Algorithm 7
      2.1.2 Rolling 6 9
      2.1.3 Other rule based designs 10
   2.2 Model Based Designs 10
      2.2.1 Continual Reassessment Method (CRM) 12
      2.2.2 Escalation With Overdose Control (EWOC) 19
      2.2.3 Bayesian Decision Procedure applied to a Logistic Regression Model .22
   2.3 Overview 26
3. Investigating data trends in Phase I trials 29
   3.1 38 Phase I Trials in Molecularly Targeted Agents 29
   3.2 Data cleaning 30
   3.3 Occurrence of toxicities 32
   3.4 Covariate EDA 35
      3.4.1 Age 35
      3.4.2 Gender 37
      3.4.3 Primary Tumours 39
      3.4.4 Interaction between Gender/Age and Primary Tumour Type 43
      3.4.5 Family of Toxicity 43
4. Methodology – ICSDP 46
   4.1 The Interval-Censored Survival Model 46
   4.2 Prior Information 52
   4.3 Gain Function 54
   4.4 Escalation Features 55
   4.5 Proportional Odds Decision Procedure 58
5. The Interval-Censored Survival Decision Procedure: A Simulation Study 61
   5.1 A Comparison of 3 Designs 61
   5.2 Data Generation Scenarios 61
      5.2.1 Introduction 61
      5.2.2 Proportional Odds Model 62

5.2.3 Interval-Censored 67
5.2.4 Proportional Odds with dose as covariate 70
5.2.5 Differences 72
5.3 Pseudo-data Prior Information 76
5.4 Escalation Procedure 79
5.5 Results 82
5.5.1 Generated by the Proportional Odds Model with log(dose) 82
5.5.2 Generated by the Interval-Censored Survival Model 87
5.5.3 Generated by the Proportional Odds Model with dose 91
5.6 Discussion and Investigation of Results 95
5.7 Continuing the procedures to the maximum number of cohorts 96
5.8 Considering the effect of censoring 98
5.8.1 Results for Non-Informative Censoring 102
5.8.2 Results for Informative Censoring 103
5.9 The effect of incorrect assumptions in the trial design 103
5.9.1 Investigating a TD33 105
5.9.2 Investigating a TD38.6 106
5.10 Conclusions 108
6. Simple variations on the ICSDP 111
6.1 Investigating the use of the variance gain function 111
6.1.1 Generation by the Proportional Odds model with log(dose) 114
6.1.2 Generated by the Interval-Censored Survival model 119
6.1.3 Variations of the variance gain 122
6.1.4 Conclusions 130
6.2 Incorporating intra-patient adjustments 131
6.2.1 The ICSDP incorporating adjustments 132
6.2.2 Results 134
6.2.3 Investigation of Results 135
6.2.4 Imposing restrictions on the intra-patient adjustments 142
6.2.5 Conclusions 145
7. Including Baseline Covariates into the Escalation Procedure 147
7.1. Introduction 147
7.2. Data Generation Methods 148

7.2.1 Covariates 148
7.2.2 PO Model 151
7.2.3 ICS Model 154
7.3 Pseudo-data 155
7.4 Escalation Procedure 157
7.5 ICSDP with baseline covariates 163
7.5.1 Results 163
7.5.2 Removing the pseudo-data from the final analysis 168
7.6 Investigating the amount of Pseudo-data to use 171
7.6.1 Results from n=3/4 pseudo-observations per covariate category per dose 171
7.6.2 Results from n=1 pseudo-observations per covariate category per dose 175
7.6.3 Results from n=1.5 per covariate category per dose 177
7.6.4 Conclusions 179
7.7 Investigating the Inclusion of a Prior Covariate Effect 179
7.7.1 Results 182
7.7.2 Results with an increased amount of pseudo-data 184
7.7.3 Conclusions 185
7.8 Overall Conclusions 185
7.9 Remarks 187
8. Allowing For Lower Grade Toxicities in the Analysis of DLTs 189
8.1 Motivation 189
8.2 Methodology 193
8.3 Simulation Methods 194
8.4 Escalation Procedure 202
8.5 Scenarios 207
8.6 Results 211
8.6.1 ICSDP incorporating LG toxicities 212
8.6.2 Testing the procedure when there are contradictions between data generation and analysis 218
8.6.3 Making the data more realistic and further testing of assumptions 222
8.7 Conclusions 227
9. Conclusions and Remarks 230
9.1. Overall Conclusions 230
9.2. Extensions and Further Work 233

10. Appendices 236

Appendix 1: Rearranging the complementary log-log link function in terms of the TD for *s* cycles. 236

Appendix 2: Deriving the asymptotic variance for the complementary log-log link function for 3 cycles 237

Appendix 3: Inverted Information Matrix 245

Appendix 4: Extending the asymptotic variance for the complementary log-log link function to *s* cycles 246

Appendix 5: Deriving the asymptotic variance for the complementary log-log link function for 3 cycles with 2 covariates 248

11. Bibliography 257

List of Figures

Figure 5- 1: True Dose-Response Relationship used for Simulation 63
Figure 5- 2: Data simulated by PO model with log dose for 3 cycles – – – , analysed by ICS model for 3 cycles ——..... 73
Figure 5- 3: Data simulated by ICS model for 3 cycles ——, analysed by LR model with log dose for 3 cycles – – –. 73
Figure 5- 4: Data simulated by ICS model for 1 cycle —— and analysed by LR for 1 cycle – – –. 74
Figure 5- 5: Data simulated by PO model with dose for 3 cycles - - -, analysed by ICS model —— and LR with log dose – – – for 3 cycles. 74
Figure 5- 6: Data simulated by PO model with dose - - - for 1 cycle, analysed by LR for 1 cycle – – –. 75
Figure 5- 7: Distribution of recommended doses for different procedures when the data is generated by the PO model with log(dose), a) LRDP1, b) LRDP3, c) ICSDP 85
Figure 5- 8: Distribution of doses recommended for different procedures when the data is generated by the ICS model, a) LRDP1, b) LRDP3, c) ICSDP. 89
Figure 5- 9: Distribution of recommended doses for different procedures when the data is generated by the PO model with dose, a) LRDP1, b) LRDP3, c) ICSDP..... 93

Figure 6- 1: Distribution of recommended doses for different procedures when the data is generated by the PO model with log(dose), a) LRDP1, b) LRDP3, c) ICSDP. 117
Figure 6- 2: Distribution of dose recommendations for different procedure when the data is generated by the ICS model, a) LRDP1, b) LRDP3, c) ICSDP. 121
Figure 6- 3: Histograms to show the distribution of the estimated TDs when simulated by PO model. 136
Figure 6- 4: Histograms to show the distribution of the estimated TDs when simulated by ICS model. 137

Figure 8- 1: Distribution of true Individual TDs. 204

List of Tables

Table 2- 1: Pseudo-data for Logistic Regression Decision Procedure. 23

Table 3- 1: Patients experiencing their first DLT^+ in each cycle. *average percentage per cycle for 10 cycles. 32
Table 3- 2: Trial 64 – Patients experiencing their first DLT^+ in each cycle. 34
Table 3- 3: Mean age of patients in each cycle for all patients, and patients with DLT^+. 35
Table 3- 4: Number of patients in each age category. 36
Table 3- 5: Numbers of patients with their first DLT^+ in each cycle for different age categories. *average percentage per cycle for 10 cycles. 36
Table 3- 6: Number of patients with a DLT^+ and a protocol specified DLT for each gender. 38
Table 3- 7: Number of patients with first DLT^+ in each cycle for each gender. *average percentage per cycle for 10 cycles. 38
Table 3- 8: Numbers of patients, and number with DLT^+ and protocol specified DLTs for each primary tumour type. 40
Table 3- 9: Number of patients with first DLT^+ in each cycle of occurrence for each primary tumour type. *average percentage per cycle for 10 cycles. 41
Table 3- 10: Number of patients with a first DLT^+ in each cycle for each type of toxicity. 44

Table 4- 1: Pseudo-data for $d_{(1)}$ and $d_{(k)}$, conditional on previous cycles of therapy. . 53

Table 5- 1: Checking the simulation method does not match the analysis method. 67
Table 5- 2: Checking the simulation method does not match the analysis method. 69
Table 5- 3: Checking the simulation method does not match the analysis methods. .. 71
Table 5- 4: Pseudo-data for the LRDP1 77
Table 5- 5: Pseudo-data for the LRDP3 77
Table 5- 6: Pseudo-data for the ICSDP 79
Table 5- 7: Results from 1000 trials, generated by the PO model and escalated with the patient gain. TD=366mg/m^2 82
Table 5- 8: Results from 1000 trials, generated by the ICS model and escalated with the patient gain. TD=366mg/m^2 87
Table 5- 9: Results from 1000 trials, simulated by the PO model with dose and escalated with the patient gain. TD=366mg/m^2. 91
Table 5- 10: Results from 1000 simulations by the ICS model with the patient gain and a fixed number of patients (20 cohorts of size 3). TD=366mg/m^2 97
Table 5- 11: Effect of non-informative censoring 99
Table 5- 12: Effect on informative censoring. 101

Table 5- 13: Results from 1000 trials, simulated by the ICS model and escalated with the patient gain with 10% non-informative censoring. TD=366mg/m^2..... 102
Table 5- 14: Results from 1000 trials, simulated by the ICS model and escalated with the patient gain with 10% informative censoring. TD=366mg/m^2..... 103
Table 5- 15: Possible probability differences 104
Table 5- 16: Pseudo-data for a procedure looking for the TD33..... 105
Table 5- 17: 100 simulations investigating a TD33. TD=380mg/m^2..... 106
Table 5- 18: Pseudo-data for procedure looking for TD38.6..... 107
Table 5- 19: 100 simulations investigating a TD38.6. TD=435mg/m^2..... 107
Table 5- 20: 100 simulations investigating a TD38.6 with safety stopping rule p=0.5. TD=435mg/m^2..... 108

Table 6- 1: Results from 100 trials, simulated by the PO model and escalated with the variance gain. TD=366mg/m^2..... 114
Table 6- 2: Results from 100 trials simulated by the PO model and escalated with the variance gain without trials stopped for safety. TD=366mg/m^2..... 115
Table 6- 3: Results from 100 trials, simulated by the PO model and escalated with the variance gain for permissible doses. TD=366mg/m^2..... 116
Table 6- 4: Results from 100 trials, simulated by the ICS model and escalated with the variance gain. TD=366mg/m^2..... 119
Table 6- 5: Results from 100 trials, simulated by the ICS model and escalated with the variance gain without trials stopped by the safety rule..... 119
Table 6- 6: Results from 100 trials, simulated by the ICS model and escalated with the variance gain for permissible doses. TD=366mg/m^2..... 120
Table 6- 7: Results from 100 trials, simulated by the PO model and escalated with the unrestricted variance gain. TD=366mg/m^2..... 123
Table 6- 8: Results from 100 trials, simulated by the ICS model and escalated with the unrestricted variance gain. TD=366mg/m^2..... 123
Table 6- 9: Dose administrations recommended for cohort 2 when escalated by the unrestricted variance gain and simulated by the ICS model..... 124
Table 6- 10: Administration of doses in one simulated trial (escalated with the unrestricted variance gain and data generated by the ICS model) when the safety rule is not used, excluding cohort 1 and pseudo-data. 125
Table 6- 11: Results from 100 trials, simulated by the PO model and escalated with the variance gain for permissible doses, until the same average number of cohorts as in the patient gain (Table 5-7) have been recruited. 128
Table 6- 12: Results from 100 trials, simulated by the ICS model and escalated with the variance gain for permissible doses until the same average number of cohorts as in the patient gain (Table 5-8) have been recruited..... 129
Table 6- 13: Results from 1000 trials simulated by PO model or ICS model, escalated with patient gain, allowing intra-patient adjustments and not. TD=366mg/m^2..... 134
Table 6- 14: One trial, data simulated by ICS model, escalated with intra-patient adjustments. 139

Table 6- 15: Comparing Intra-Patient Escalation to No Intra-Patient Escalation..... 140
Table 6- 16: Results from 100 trials simulated by PO model and ICS model, escalated with patient gain, allowing intra-patient escalation until cohort 4 begins. 144
Table 6- 17: Results from 100 trials simulated by PO model and ICS model, escalated with patient gain, allowing intra-patient escalation after cohort 5 begins. 145

Table 7- 1: Probability of DLT (rounded to 2dp) and TD31.6 for each category in cycle 1. 151
Table 7- 2: Parameter values when covariates are coded -0.5, 0.5. 152
Table 7- 3: Cumulative probabilities of DLT for each subgroup for each cycle. 152
Table 7- 4: Parameter values when covariates are coded 0, 1. 153
Table 7- 5: Parameter values when covariate values are coded -0.5, 0.5. 154
Table 7- 6: Conditional probabilities of DLT for each subgroup for each cycle. 154
Table 7- 7: Parameter values when covariate values are coded 0, 1 155
Table 7- 8: Probabilities of DLTs associated with different doses and dose levels from ICS data simulation model. 156
Table 7- 9. Pseudo-data for all categories of patients 156
Table 7- 10: Results from 100 trials simulated by ICS model with covariate values 0, 1 or -0.5, 0.5. 164
Table 7- 10 cont.: Results from 100 trials simulated by ICS model with covariate values 0, 1 or -0.5, 0.5 165
Table 7- 11: Results from 100 trials simulated by PO model with covariate values 0, 1 or -0.5, 0.5. 165
Table 7- 12: Some results from Tables 7-8 and 7-9 reanalysed without the pseudo-data. 169
Table 7- 13: Results from 100 trials simulated by ICS or PO models with n=3/4 per category for pseudo-data. 172
Table 7- 14: Results from 93 trials when generated by ICS, 98 when generated by PO. With n=3/4 per subgroup per dose for pseudo-data without trials that stopped for safety. 173
Table 7- 15: Results from 100 trials simulated by ICS or PO models with n=1 per subgroup per dose for pseudo-data. 175
Table 7- 16: Results from 100 trials simulated by ICS or PO models with n=1.5 per subgroup per dose for pseudo-data. 177
Table 7- 17: P(DLT) at various doses for each covariate category 180
Table 7- 18: Parameter values associated with the pseudo-data 181
Table 7- 19: Pseudo-data for all subgroups of patients with a prior covariate effect 181
Table 7- 20: Results from 100 trials simulated by ICS or PO models with covariate values 0, 1 with n=1.5 per subgroup per dose with a prior covariate effect for pseudo-data. 183
Table 7- 21: Results from 100 trials simulated by ICS or PO models with covariate values 0, 1 with n=3 per subgroup per dose and a prior covariate effect for pseudo-data. 184

Table 7- 22: Results from 1000 trials simulated by ICS or PO models with n=1.5 for pseudo-data. 186

Table 8- 1: Number of patients with their first DLT in each cycle..... 191
Table 8- 2: Numbers of LGTs for patients in cycles prior to the cycle of interest. For patients that did and did not experience a DLT in the specified cycle. 191
Table 8- 3: Numbers of previous LGTs for patients that did and did not experience a DLT in a given cycle. Including LGTs that occurred in the same cycle. 192
Table 8- 4: Occurrence of LGTs in each cycle. 193
Table 8- 5: Mean number of LGTs for each dose. 195
Table 8- 6: Parameter values from model (8.1) used for simulation. 198
Table 8- 7: Parameter values used for generation when an interaction is included as in model (8.2). 199
Table 8- 8: Pseudo-data for the implementation of LGTs into the initiation of the ICSDP with no LG*dose interaction. 206
Table 8- 9: Pseudo-data for the implementation of LGTs into the initiation of the ICSDP with LGT*dose interaction. 207
Table 8- 10: Scenarios for investigation. 209
Table 8- 11: Results from Scenario 1, 1000 trials simulated by the ICS model. 213
Table 8- 12: Mean Estimates associated with individual target doses after 1000 trials from Scenario 1. 214
Table 8- 13: Differences in mg/m^2 and P(DLT) between adjacent dose levels. 215
Table 8- 14: Results from Scenario 2, 100 trials simulated by the ICS model. 216
Table 8- 15: Mean Estimates associated with individual target doses after 100 trials from Scenario 2. 217
Table 8- 16: Results from Scenario 3, 1000 trials simulated by the ICS model. 218
Table 8- 17: Mean Estimates associated with individual target doses after 1000 trials from Scenario 3. 219
Table 8- 18: Results from Scenario 4, 1000 trials simulated by the ICS model. 220
Table 8- 19: Mean Estimates associated with individual target doses after 1000 trials from Scenario 4. 221
Table 8- 20: Results from Scenario 5, 1000 trials simulated by the ICS model. 222
Table 8- 21: Mean Estimates associated with individual target doses after 1000 trials from Scenario 5. 223
Table 8- 22: Results from Scenario 6, 1000 trials simulated by the ICS model. 224
Table 8- 23: Mean Estimates associated with individual target doses after 1000 trials from Scenario 6. 224
Table 8- 24: Results from Scenario 7, 1000 trials simulated by the ICS model. 225
Table 8- 25: Mean Estimates associated with individual target doses after 1000 trials from Scenario 7. 226

Glossary

B&G	Breast & Gynaecological
CDF	Cumulative Distribution Function
CI	Credible Interval
CNS	Central Nervous System
CRM	Continual Reassessment Method
CRML	Continual Reassessment Maximum Likelihood
DLT	Dose Limiting Toxicity
EWOC	Escalation With Overdose Control
GI	Gastrointestinal
Gn	Grade n toxicity (n=1,2,3)
HR	Hazard Ratio
ICS	Interval-Censored Survival
ICSDP	ICS Decision Procedure
LGT	Lower Grade Toxicity
LR	Logistic Regression
LRDP	LR Decision Procedure
MLE	Maximum Likelihood Estimate
MTA	Molecularly Targeted Agent
MTD	Maximum Tolerated Dose
OR	Odds Ratio
PD	Pharmacodynamics
P(DLT)	Probability of DLT
PFS	Progression Free Survival
PK	Pharmacokinetics
TD	Target Dose
TD_{TTL}	TD (corresponding to) TTL
TD20	TD (corresponding to) 20% (toxicity level)
TD31.6	TD (corresponding to) 31.6% (toxicity level)
TD50	TD (corresponding to) 50% (toxicity level)
TITE-CRM	Time To Event-CRM
TTD	True TD
TTL	Target Toxicity Level

1. Introduction

Dose-escalation studies are carried out in Phase I of the drug development process as the start of the clinical phase. The objective of these studies is to determine one or a few dose levels of a new drug which are deemed safe enough to carry forward to Phase II trials, where the efficacy of the drug is investigated. The definition of a safe dose is one that does not exceed a certain safety level, usually defined as a proportion of patients experiencing some level of toxicity, which can be translated to the probability of a patient experiencing a toxic event. In particular, doses corresponding to specific levels of safety are sought in order to focus later studies to specific dose ranges. These doses are called target doses (TD). One particular TD is the maximum tolerated dose (MTD), which is defined as the dose that causes the maximum tolerable level of toxicity in subjects. Toxicity is defined as an adverse event that is experienced alongside taking the new drug. There are different types of toxicity for drugs associated with different therapeutic areas such as renal, cutaneous, gastro-intestinal etc. There are also different levels of toxicity that can occur. These range from grades 1-5, 1 being mild toxicity and 5 being fatal. The specific grade of toxicity is relative to the drug in question and the disease being targeted. For most relatively non-toxic drugs, the interest lies in investigating very mild toxicities, so the dose-finding studies can begin with healthy volunteers. However, when investigating particularly toxic treatments such as cytotoxic treatments in cancer therapy which target rapidly developing cells (such as tumour cells), it is assumed that there is an association between toxicity and efficacy. Therefore some tolerated level of toxicity is required to ensure the drug is working effectively, and it is simply not ethical to use healthy volunteers. It would also not be informative to use healthy volunteers since the desired effect of the drug (the effect on reducing tumour cell reproduction) would not be

observable since the patients do not have the disease in the first place. The patients used for trials of these types of treatments would therefore be patients with the disease already, such as cancer patients. In other settings it may be more feasible to have healthy volunteers where the effect of the drug can be observed by changing levels of biomarker cells or hormones but if tumour cells are required to observe efficacy, only cancer patients themselves should be utilised.

Along with the main aim of establishing a safe dose, the safety profile of the drug is also investigated in Phase I. The pharmacokinetics (PK) of the drug (that is, how the drug moves around and through the body) are investigated here. This is often done by looking at how the concentration of the drug in the body changes with time. The pharmacodynamics (PD) of the drug are also investigated here (the effect the drug has on the body) by looking at receptors in the body and the effect of the drug on these receptors. These properties can even be used to initiate dose-finding procedures by using an appropriate dose-response model to predict a very safe dose for the initial dose.

The general idea of the dose-escalation procedure is to begin the study by allocating a dose to patients that is believed to be safe. This is often the lowest dose level available from a set of admissible doses determined by pre-clinical investigation. A pre-defined level of toxicity (Target Toxicity Level, TTL) is decided and a safe dose is allocated to the first cohort of patients. The cohort is followed up to see whether Dose Limiting Toxicities (DLTs) occur. A DLT according to the specifications laid out in the study specific protocol is generally a grade 3/4 toxicity which if observed leads to the current treatment being stopped (possibly adjusted or withdrawn). DLTs are defined in the study protocol to be disease specific. The definition of a DLT may not just be an occurrence of a grade 3/4 toxic event, but might include multiple occurrences of lower

grade toxicities (LGTs). Based on the data from the first cohort, the dose to be allocated to the next cohort is determined and administered. This procedure continues until some stopping criterion has been reached. The stopping criteria typically consist of a safety criterion, a precision criterion and a maximum number rule. The safety rule stops the procedure if the recommended dose is associated with a probability of observing a DLT which is too high. The precision criterion stops the procedure when enough information (by way of DLTs) has been observed to conclude with sufficient confidence that a certain dose is the dose that corresponds to the TTL. This dose might be the MTD or it might be another TD that is deemed a suitable compromise between efficacy and safety.

Dose-escalation studies for cancer treatments are not quite as straightforward as for other treatments. First, the patient population must be cancer patients, as it is not ethical or useful to use healthy volunteers. The TTL is also quite difficult to establish as there are different grades and types of toxicity that can occur for a cancer patient. Toxicities which are deemed 'dose-limiting' (too high to warrant escalating to a higher dose) are defined and a TTL associated with these grades of events needs to be decided. The TTL that is decided is based on the assumption that in order to be effective, cytotoxic treatments are expected to cause some toxicity. Therefore the TTL is a compromise between safety and presumed efficacy. Discussion will also be required regarding the definition of what is considered dose-limiting. For example, whether it is just the occurrence of grade 3/4 toxicities or if it includes occurrences of LGTs. Grade 5 toxicities (death) may be included as dose-limiting but they may also be censored from the analysis. This would be defined in the protocol,. In these trials the initial dose for the allocation is usually the lowest dose, unless there is a lot of previous information about this drug (maybe from preclinical data). The escalation is

then carried out under the same general algorithm as described previously. Patients are observed for DLTs, doses to be allocated to subsequent patients are decided based on data so far, and the trial continues until a stopping criterion is met.

Another aspect of cancer dose-escalation studies that differs from studies in general is that the treatment is normally administered in cycles of therapy with periods of no treatment between cycles. This prolongs the trial, so it is usually only the first cycle of therapy that is included in the escalation process. If the procedure waited for every patient to complete all cycles of therapy, it could be an unfeasibly long time before the next patient received their selected dose and the trial could last an impractically long time. Obviously there are some issues with this since a lot of information (particularly if there are many cycles of therapy) is being disregarded, and in the case of newer Molecularly Targeted Agents (MTAs), later toxicities may still need to be included (Postel-Vinay et.al. [1]).

Dose-finding studies for cancer therapies have been the source of investigation for a long period of time. Traditional rule-based designs, such as the 3+3 algorithm [2], are widely recognised and used. With these designs, doses are escalated sequentially until DLTs are observed and then a set of rules dictate how to proceed. They are, however, largely inefficient and usually produce estimates of recommended doses for later phases of investigation that are sub-therapeutic. Model-based designs have been developed which use a parametric model to describe the relationship between dose and response, where response is usually the probability of observing a DLT. Within these designs a set of stopping criteria as mentioned already, are implemented to stop the procedure either when the recommended TD is estimated with sufficient precision, a safety rule is breached or a maximum number of cohorts have been recruited.

Existing designs however have traditionally only utilised information on the occurrence of DLTs from the first cycle of therapy. All patients are treated at the same dose and differences in patients' tolerabilities, which might be related to patient specific covariates or tolerance markers are not considered. Chapter 2 presents the findings of a literature review on existing designs used in practice, and discusses the benefits and weaknesses of the different approaches.

This thesis develops a new procedure to incorporate data from later treatment cycles, by using an Interval-Censored Survival (ICS) model within a Bayesian Decision framework. The ICS decision procedure (ICSDP) approach involves looking at the occurrence of a patient's first DLT. It models the probability that the first DLT occurs in each specific cycle via the probability of a DLT during that specific cycle, conditional on having no DLT in any previous cycle. In doing this the conditional properties of the ICS model allow multiple cycles of therapy to be used for analysis. This procedure therefore allows patients to contribute to the analysis for every cycle of therapy they have completed until the first occurrence of a DLT. Chapter 3 discusses the motivation for adopting this method by conducting some exploratory data analysis on completed trials. Chapter 4 presents the methodology associated with the ICS model as well as aspects of the Bayesian Decision procedure, such as the choice of prior information and how to implement it, the calculation of the posterior distribution and the derivation of the stopping criteria. Chapter 5 investigates the advantages offered by this method by comparing this procedure to other existing procedures which incorporates just one cycle of therapy [4]. The procedure used for comparison is the logistic regression decision procedure (LRDP) which is described in detail in Chapter 2. Both are then also compared to a LRDP which considers multiple cycles of therapy as one fixed period of observation. This is a compromise to be

considered when considering incorporating later cycles of therapy easily. The new ICSDP can also allow patients to change doses between cycles if the dose level that is believed to be closest to the estimated TD changes as a new cohort is recruited. This use of intra-patient dose adjustments is considered in Chapter 6.

The idea of personalised escalation procedures is an attractive idea and one that is becoming more popular in practice, one example is discussed in Babb and Rogatko [5]. The use of baseline covariates within the ICS model is a way to allow different categories of patients to be recommended different target doses. This is considered in Chapter 7. The adaptation of the existing procedure to allow for different categories of patients in the initiation of the escalation and the decision making process is discussed, along with some of the implications of allowing a range of doses to be recommended at the end of the trial.

Furthermore, an extension of using patient specific baseline characteristics is to allow covariates that reflect a patient's reaction to the drug. Time-changing covariates are considered, such as the occurrence of LGTs, which act as a marker or indicator for an increased chance of a DLT. The chance of experiencing LGTs is likely to change dependent on the length of time in the study. Furthermore, the increase or reduction in prevalence of LGTs can be used to allow dose changes between cycles in order to ensure patients receive doses targeted to their specific tolerabilities. Issues that arise from allowing every patient to be recommended a different dose are considered and investigated in Chapter 8.

The results from all of the investigative Chapters (5, 6, 7 and 8) are summarised and reviewed in Chapter 9 and some conclusions and recommendations are offered for the design and conduct of Phase I dose-finding studies for cancer therapies.

2. Literature Review of Existing Phase I Dose-Finding Procedures

Dose-finding procedures can be split into two different categories. The first focuses on using a system of rules to make escalation decisions, the so called rule-based designs. These designs are very simple to implement and are therefore very common. The second category assigns a parametric model to the relationship between dose and the probability of observing a DLT for a randomly chosen patient from the population (P(DLT)), and uses the model to analyse the observations to obtain predictions for which dose corresponds to the required TTL. These are therefore referred to as model-based designs.

This chapter looks in detail at some common procedures in both the rule-based and model-based categories. The methods and conduct of each procedure are described and some discussion of the benefits and shortcomings of these procedures are given.

2.1 Traditional Rule Based Designs

Rule Based Designs are often used for dose-finding in practice and are very popular because of the simplicity of the procedure. No models are used in the implementation, although the assumption that the probability of DLT increases with dose is still made, and it is very easy for clinicians to understand the logic.

2.1.1 3+3 Algorithm

The 3+3 algorithm [2] is a very simple procedure which is based on the idea that if a dose produces no more than an observed rate of DLTs of 33%, then it is the MTD. The conduct of the trial is summarized by Berry et. al. [3] as follows:

1. A set of dose levels (in ascending order) is decided,

2. The lowest dose level is administered to 3 patients,
    a. If 0/3 patients experience a DLT, the procedure escalates one dose level for the next cohort of 3 patients.
    b. If 1/3 patients experience a DLT, the next cohort of 3 patients is treated at the same dose level.
    c. If 2/3 or 3/3 patients experience a DLT, the trial is aborted with no safe dose.
3. If the procedure escalates one dose level (as in step 2a), Step 2 is repeated with the next dose.
    a. If 0/3 patients experience a DLT on the new dose, the procedure escalates one dose level for the next cohort of 3 patients.
    b. If 1/3 patients experience a DLT, the next cohort of 3 patients is treated at the same dose.
    c. If 2/3 or 3/3 patients experience a DLT on the new dose, the trial is stopped and the dose below the current dose is classed as the MTD.
4. If the procedure has repeated the same level in successive cohorts (2b or 3b);
    a. If 0/3 patients experience a DLT, the procedure escalates one dose level for the next cohort of 3 patients.
    b. If 1/3 patients experience a DLT on the repeated dose such that 2/6 patients have experienced a DLT on that dose level, the next cohort is treated at the preceding dose level providing that only 3 patients have been treated at it. If 6 patients have been treated at the preceding dose level the trial is stopped and that lower dose is declared as the MTD.
    c. If 2/3 or 3/3 patients experience a DLT, the trial is stopped and the previous dose is classed as the MTD.

5. Continue until an MTD has been established.

This procedure tries to ensure that the dose classed as the MTD does not produce a probability of DLT greater than 0.33 (2/6) even though the true toxicity rate is not necessarily known.

While being cautious in design, the 3+3 algorithm is extremely inefficient, particularly when there are many dose levels and the starting dose (first dose level) is much lower than the true MTD. In the case of cancer treatments, the starting dose is nearly always the lowest dose, so it can be assumed that it is far below the true MTD (especially if there are many possible dose levels) and many patients will be treated at sub-therapeutic doses.

2.1.2 Rolling 6

The Rolling Six design is an extension of the 3+3 algorithm developed by Skolnik et. al. [6], for the purpose of shortening the duration of pediatric Phase I trials. This method was developed after extensive investigation into previous pediatric Phase I trials conducted by Lee et. al. [7] suggested that on average 5.1 patients were treated at each dose level. The general concept remains the same as for the 3+3 algorithm, with dose escalation occurring when no DLTs have been observed in the three patients at a specific dose level. If one of the three experiences a DLT, the dose is repeated for another cohort of three patients. If two of the three experience a DLT, the dose is de-escalated for the next cohort. The difference between the Rolling Six and the 3+3 is that patients are continually accrued and suspension of accrual occurs after every six patients as opposed to after every three patients. Therefore, two cohorts are observable in any one observation period, possibly on different doses if the first cohort experienced no DLTs or two DLTs. Since accrual is continuous, one patient is recruited at a time, so if a patient in a cohort is not evaluable, a new patient can be

added to this cohort to complete the cohort of three patients. This will then allow quicker evaluation of dose levels. The simulation study conducted by Skolnik et. al [6] showed that the Rolling Six shortened the length of Phase I trials in every scenario simulated, compared to the 3+3 design, due to the smaller number of times the trial is suspended between cohorts (every 6 patient rather than every 3).

Although the Rolling 6 method shortens the timeline of Phase I trials, the general method remains the same as for the 3+3. This suggests the method is very inefficient and many patients could still be treated at sub-therapeutic doses. In particular, the results shown in [6] suggest that there is a slight increase in the number of patients required to complete the trial. This is unethical, especially when many of these patients are treated at sub-therapeutic doses.

2.1.3 Other rule based designs

The other rule based designs tend to expand upon the traditional 3+3 (e.g. 2+4 etc.) and add extra steps to the procedure. For example, the initial dose could be calculated from PK and PD data for each patient and then escalated one dose level at a time. If DLTs start occurring then the rules associated with the 3+3 method can be incorporated. Another version is Accelerated Titration [8], where the 3+3 method is used but intra-patient (within a patient but between cycles of therapy) escalation can also occur. This involves allowing patients to change dose levels between successive cycles of therapy according to the accruing information from the overall study however the information from later cycles of therapy and potentially different doses may not be utilised in the analysis.

2.2 Model Based Designs

An alternative to a rule-based design is a model-based approach, which can be used to relate the probability of DLT to dose. This typically assumes an increasing probability

of DLT with dose. The decisions (or rules) for escalating/de-escalating doses are made via the use of gain functions.

A Bayesian procedure is often used which relies upon prior information to initiate the escalation. The prior information is updated with every observation to produce posterior information about the parameters of the model, and in turn the dose corresponding to the TTL. This is called the Bayesian decision procedure (Whitehead, Brunier [9]) and it consists of five main components.

1. A parametric model is chosen to represent a monotonically increasing relationship between dose and probability of DLT.
2. A prior distribution is assigned to the parameter(s) of the model.
3. Once data have been obtained, analysis of the observed data and the prior information produces a posterior distribution for the parameter(s) of the model.
4. A set of possible actions is defined. In this case the actions relate to the choice of one of a set of possible dose levels ($d_{(j)}$, $j=1....k$) to allocate to the next cohort.
5. A gain function is required in order to choose between the actions.

There are maximum likelihood versions of these procedures which do not involve the use of prior information, resulting in just four steps. These are: choosing a model; analysing the observed data to obtain maximum likelihood estimates of the parameter(s); creating a set of possible actions and creating a gain function to decide between the actions. Such procedures start the escalation at the lowest possible dose and the doses are escalated one level at a time until a DLT is observed. Once there is heterogeneity in the observations of DLTs (i.e. both occurrences and non-occurrences

of DLTs), the likelihood can be constructed and maximised to obtain maximum likelihood estimates for the parameters, from which the TD associated for the relevant TTL can be calculated. There are different approaches to the model based designs with different models, different endpoints or different gain functions.

2.2.1 Continual Reassessment Method (CRM)

The Continual Reassessment Method (CRM) was developed in the early 1990s by O'Quigley et al. [10], and was one of the first Bayesian model based approaches to dose-finding.

For a set of discrete dose levels $d_{(j)}, j = 1,\ldots,k$, the corresponding probability of DLT can be defined as $p_{(j)}, j = 1,...,k$. Interest then lies in finding the dose that corresponds to a pre-specified TTL. This dose will be the TD_{TTL}.

As described in the previous section, there are five components to a Bayesian Decision Procedure.

The first is the model chosen to represent the dose-response relationship. For the CRM this is a one parameter model which is defined as:

$$p_{(j)} = f(x_{(j)}, \beta),$$

where $x_{(j)}$ is the function of the dose level $d_{(j)}$ given by:

$$x_{(j)} = \zeta\left(d_{(j)}\right).$$

The function of the dose level can simply be a transformation of the dose, e.g. a log transformation. The function f can be any suitable model that depicts the dose-response relationship. An example as discussed in [10] is the hyperbolic tangent function:

$$p_{(j)} = \left(\tanh\left(x_{(j)}\right)\right)^{\beta}.$$

The second step in the Bayesian Decision Procedure is to place a prior distribution on the parameter β which is defined with a density function $g^0(\beta)$. The distribution associated with this density function is often a Normal distribution with mean parameter β^0 and variance parameter $\sigma_\beta{}^2$. O'Quigley [10] uses an $Exp(1)$ distribution.

The procedure begins by creating initial guesses $p^0{}_{(j)}$ for the probability of toxicity, $p_{(j)}$, associated with the different dose levels $j = 1,...,k$, based on the prior distribution of β. $g^0(\beta)$ is usually fixed so that the prior probabilities are strictly increasing with dose, i.e. $p_1{}^0 < p_2{}^0 < ... < p_k{}^0$ and the lowest dose corresponds to the TTL. The dose with $p^0{}_{(j)}$ closest to the TTL (the lowest dose) is administered to the first cohort.

The third step of the Bayesian procedure is to compute a posterior distribution utilising information up to and including patient i for the parameter β. The Bayes estimate of the parameter β_i is calculated by the following equation:

$$\beta_i = \frac{\int_{-\infty}^{\infty} \beta L_i(\beta) dG^0(\beta)}{\int_{-\infty}^{\infty} L_i(\beta) dG^0(\beta)}.$$

Here, the derivative $\frac{dG^0(\beta)}{d\beta}$ is the density function $g^0(\beta)$ as described previously, and $L_i(\beta)$ is the Binomial likelihood given by:

$$L_i(\beta) = \prod_{h=1}^{i} \{f(x_h, \beta)\}^{Y_h} \{1 - f(x_h, \beta)\}^{1-Y_h} \qquad (2.1)$$

$Y_h = 1$ if patient h has a DLT and 0 if not where $h = 1,...,i$. If patient h receives dose d_h, then $x_h = \zeta(d_h)$.

For the hyperbolic tangent function, or indeed any function of the form $p_{(J)} = \left(a\left(x_{(J)}\right)\right)^{\beta}$:

$$p^{i}_{(J)} = \left\{p^{0}{}_{(J)}\right\}^{\beta_i / \beta^0}.$$

With this formulation, the values of $x_{(J)}$ are no longer required to conduct the procedure. The dose level $d_{(J)}$ that has $p^{i}_{(J)}$ closest to the TTL, is administered to the next cohort.

The fourth and fifth steps of the procedure are to choose a set of actions, and a gain function to select the appropriate action. The set of actions are, as usual in model-based designs, the discrete set of dose-levels selected for administration. The gain function in this particular setting is to choose the dose $d_{(j)}$ that minimises the distance between the P(DLT) at that dose level j and the TTL. P(DLT) is calculated by substituting into f the Bayes estimate β_i (incorporating information from up to and including patient i) and the dose $d_{(J)}$.

The procedure will continue until some pre-specified safety criterion is breached or a precision criterion has been achieved. If neither of these stopping rules are implemented, the procedure will continue until the pre-determined sample size has been obtained. The final number of patients recruited is recorded as N and the MTD is then the dose that would have been administered to patient $N+1$.

The issue with the basic CRM is that it can be quite unconservative. Although the prior distribution for the parameter β can be fixed to administer the lowest available dose to the first cohort, after the first few observations the dose which seems to produce the number of toxicities closest to the TTL is chosen for administration to the next cohort. This could result in skipping many dose levels due to underestimating probabilities of toxicity based on few observations. A response to this is the two stage CRM which begins as a sequential dose allocation and only switches to the model based CRM once a DLT is observed. The sequence of doses is sorted into ascending order of probability of toxicity and the dose which corresponds to the lowest probability is administered first (as in most dose-finding procedures). If no DLT is observed at this dose, the next dose level up is administered to the next cohort. This continues for as long as there are no DLTs. Once a DLT occurs, the procedure switches to the model based method, where observations from each of the dose levels are analysed and a posterior distribution for the parameter is determined. This posterior distribution after i patients is used to produce an estimate for the parameter which, when used with the different dose-levels in the model $p_{i(j)} = f(x_{(j)}, \beta_i)$, is used to find the dose that produces a probability of DLT closest to the TTL. The two stage design is a much slower and more conservative escalation and the issue with this is that patients are likely to be given sub-therapeutic doses early on in the trial.

Another version of the CRM is the CRML [11] which is the frequentist version of the CRM and is based on Maximum Likelihood. No prior beliefs are used with the CRML and doses are allocated solely on the outcomes of previous patients. This eliminates the subjectivity of using prior belief created by the physician, and patients are treated based completely on responses of other patients. The issue with this method is that both DLTs and non-DLTs need to be observed before the model can be fitted. This

method is therefore often used in the two stage CRM, as once a DLT is observed and the allocation procedure switches to a model based approach, it switches to the CRML rather than the Bayesian CRM. So in the same way as for the CRM, the first dose administered to a cohort of patients, is just the lowest available dose. No prior information is used. The procedure allocates one dose level up for every cohort of patients until a DLT occurs. Once a DLT occurs, the binomial likelihood is constructed for all the observations so far. This likelihood (which is of the same form as the likelihood described in equation (2.1)) is then maximised to obtain the maximum likelihood estimate (MLE) for β. It is this MLE that is used, in place of the Bayes estimate in the CRM, in the dose-response model $p_{i(j)} = f(x_{(j)}, \beta_i)$, with each dose in the set of discrete doses. The dose which produces a probability of DLT closest to the TTL is the dose that is administered to the next cohort of patients.

All of the CRM procedures discussed so far generally only utilise information from one cycle of therapy. The responses of DLT are binary (DLT/no DLT) obtained during a fixed period of time. In cancer studies, patients receive cycles of therapy with rest periods between. If the trial were to wait for complete follow up after all cycles this could take an unfeasibly long time so often only the first cycle of therapy is observed, as it is generally believed that the majority of DLTs (if going to happen) will happen early on in the treatment. This clearly poses a problem when treatments are believed to cause late onset toxicity (maybe due to accumulation of drug in the body) or when there are many cycles of therapy (as lots of information may be disregarded). In response to this issue the Time-to-Event CRM (TITE-CRM) was created by Cheung and Chappell [12]. This method involves recruiting patients continually throughout the trial and the endpoint is whether a DLT has been observed by the end of the treatment period (all cycles). The procedure is the same as for the

standard CRM however the construction of the likelihood is slightly different. At every assessment time, patients who had experienced a toxic event contributed an event to the likelihood $\left(f\left(x_h,\beta\right)\right)$, patients who had completed all cycles of therapy without experiencing an event contributed a non-event to the likelihood $\left(1-f\left(x_h,\beta\right)\right)$. Patients who were part way through their treatment schedule without an event contributed a weighted non-event to the likelihood with the weight depending on how far through the schedule they were $(1-w_h f\left(x_h,\beta\right))$. This weight (w_h for the h^{th} patient to enter the trial) could be something relative such as the number of days in the trial/total number of days in schedule. So the likelihood for the i patients who have entered the trial is of the following form;

$$L_i(\beta)=\prod_{h=1}^{i}\left\{w_h f(x_h,\beta)\right\}^{Y_h}\left\{1-w_h f(x_h,\beta)\right\}^{1-Y_h}.$$

When a patient has an event $w_h=1$ and $Y_h=1$, so the contribution to the likelihood reduces to the same as before, $f(x_h,\beta)$. When a patient has completed therapy, $w_h=1$ and $Y_h=0$, so the contribution to the likelihood is $1-f(x_h,\beta)$. Therefore, when all patients complete therapy or have an event, the likelihood reduces to the same Binomial likelihood as in equation (2.1). When no patients have completed therapy or experienced an event so far throughout the trial, the likelihood reduces to;

$$L_i(\beta)=\prod_{h=1}^{i}\left\{1-w_h f(x_h,\beta)\right\}.$$

The motivation for this is that it is believed that the longer a patient lasts without a toxic event; the risk of them having a toxic event is reduced, so the contribution of a non-event to the likelihood should be larger for those who have lasted longer in the trial without an event than those very early on. This weighting scheme is based on the

risk of a first DLT occurring and there being no accumulation of drug in the body. The likelihood is then used after every assessment either with the prior distribution to produce posterior estimates or maximised to produce maximum likelihood estimates (MLEs) of the parameter β. The estimate is then used in the model function to find the dose with probability closest to the TTL. There are also other weight functions which can be used to change the shape of the weights with time. For example, an adaptive weighting scheme bases weights at each time on accrued observations of doses so far.

The different CRM procedures aim to solve many different issues that arise in different circumstances (e.g. treatments in different therapeutic areas) and the suggestion is that the particular CRM procedure used should be decided on a case by case basis. One main issue that arises with all procedures is the difficulty involved in computing the posterior estimates, as the Bayes Estimates of the posterior mean involve computing many complicated integrals at every assessment and this has to be done before any new patient can be allocated a dose. Although the CRML involves much easier computation, the Bayesian aspect is then lost, and possibly informative prior information is disregarded which may result in patients at the start of the trial being under or overdosed quite substantially.

The TITE-CRM is also the only version that incorporates use of later information. However extensive exploratory analysis would be needed in order to find a weighting scheme that truly represents how the drug works over. The weights are created in a way that changes depending on how far through the trial a patient is (i.e. the risk of having an event later is less than earlier, so the longer through the treatment period that a patient has survived without a DLT, the larger the weight given to the partial non-event). Cheung and Chappell [12] consider simple linear weights and adaptive

schemes, but consider the former to be overly simplistic and the latter appears to be complicated to implement. The specification of the weights, which is required before the start of the study, so that partial non-events can be included effectively from the beginning, requires knowledge of the dose-response relationship during treatment cycles, and how it changes from one cycle to the next. This information is not usually available before starting an early phase trial.

2.2.2 Escalation With Overdose Control (EWOC)

One of the main issues with the one stage CRM is the possibility of allocating patients to a dose that is higher than the true MTD. The EWOC procedure [13] aims to combat this problem by choosing a dose that (according to the current posterior belief) has a certain probability (e.g. less than 0.5) of being higher than the MTD. This is achieved by obtaining the posterior cumulative distribution function (CDF) for the MTD, and choosing a discrete dose level that most closely corresponds to a certain quantile ψ (=0.5) of the CDF, i.e.

$$\psi \geq P\left(MTD \leq d_{(J)} \mid \mathbf{y}_i\right).$$

Where $\mathbf{y}_i$ consists of the responses observed so far, up to patient i.

Having the quantile less than 0.5 (often 0.25) ensures the probability of choosing a dose for the next cohort of patients that is higher than the MTD is less than 0.5. This suggests the model is more likely to choose a dose which is less than the MTD than it is to pick one that is greater than it. Choosing ψ carefully also ensures that there is still some probability that the chosen dose is greater than the MTD, so the escalation should not be completely constrained to doses below the true MTD.

The general procedure of the EWOC design is the same as for the CRM in the sense that a model is decided for the dose-response relationship. This could again take different forms but a common one is the logistic regression model with two parameters;

$$\log\left(\frac{p_{(j)}}{1-p_{(j)}}\right)=\alpha+\beta\log\left(d_{(j)}\right),$$

Prior information is used to decide the starting dose and again, is usually fixed so that the lowest dose is chosen to be administered first. The prior information in this procedure considers the lowest dose level $d_{(1)}$ and its prior expected probability of toxicity, and also the dose that is believed to be the TD_{TTL} a priori and its relevant probability of toxicity (TTL). The parameters in the model (e.g. a logistic regression model) are reparameterised in terms of the lowest dose ($d_{(1)}$) and its believed toxicity ($p_{(1)}$), the TD_{TTL} and TTL as shown below:

$$\log\left(\frac{p_{(1)}}{1-p_{(1)}}\right)=\alpha^0+\beta^0\log\left(d_{(1)}\right),$$

$$\log\left(\frac{TTL}{1-TTL}\right)=\alpha^0+\beta^0\log\left(TD_{TTL}\right),$$

These two equations can then be solved to give prior estimates for the parameters of the model. Once the responses are observed for the first cohort of patients, the posterior cumulative distribution function (CDF) for the TD is calculated and the next dose is chosen from the set of discrete doses and according to the specified quantile ψ of the distribution. This dose level corresponding most closely to the specified quantile is administered to the next cohort, and the analysis is repeated to produce an

updated posterior CDF and a new dose for administration. This procedure continues until an estimate of the TD is produced with sufficient precision.

There is however, the issue that the doses allocated may never be greater than the true TD, or even reach the true TD, in which case the TD estimated at the end of the procedure may be a dose which is higher than any dose administered before. This is somewhat worrying as this extrapolation in estimating the TD assumes that the dose-response relationship stays the same for higher doses. If the modelled relationship were to change later on, i.e. the occurrence of late-onset toxicities increases the probability of toxicity over time, the estimation of this TD could be incorrect and the wrong dose may be carried forward to Phase II. Another issue is that very sub-therapeutic doses could be being administered. For example, if starting at the lowest dose and escalating after every set of observations, the model will identify a TD, which will most likely be smaller than the true TD due to the incorporation of the pessimistic prior information. Then the choice for the next dose will possibly be even smaller due to the probability of a lower dose being greater than the current estimated TD exceeding ψ.

Once again this procedure only uses information from the first cycle of therapy. It seems that for this procedure in particular, it may be beneficial to use information from later cycles as an accumulation of drug could show how that drug works in higher concentrations without having to give a higher dose. Therefore the extrapolation of the estimate of the TD could be more accurate as there would be some information on the drug at higher concentrations.

2.2.3 Bayesian Decision Procedure applied to a Logistic Regression Model

The simple decision procedure can be applied to a logistic regression model, which will be referred to as the Logistic Regression Decision Procedure (LRDP). In the case of the CRM, Bayes estimates are produced for the model parameters. In this procedure, Bayesian modal estimates for the parameters are produced in conjunction with the gain function to choose the doses to be administered (Whitehead and Williamson [14]). The model is shown below;

$$\log\left(\frac{p_{(j)}}{1-p_{(j)}}\right)=\alpha+\beta\log\left(d_{(j)}\right). \tag{2.2}$$

Prior distributions are usually chosen so that for safety reasons the lowest dose is given to the first cohort (typically of size three) of patients. Whitehead and Williamson [14] consider the choice of independent Beta distributions for the probability of a DLT at two different doses. This considers the minimum and maximum doses, $d_{(1)}$ and $d_{(k)}$. The Beta prior for $p_{(j)}$, j = 1, k with parameters $t^0_{(j)}$ and $u^0_{(j)}$ can be thought of as prior pseudo-data comprising $n^0_{(j)}=t^0_{(j)}+u^0_{(j)}$ observations at dose $d_{(j)}$, where $t^0_{(j)}$ is the number of patients with a DLT and $u^0_{(j)}$ is the number of patients who do not experience a DLT. The expected value of the Beta distribution with parameters $t^0_{(j)}$ and $u^0_{(j)}$ is $p^0_{(j)}=t^0_{(j)}\Big/\left(t^0_{(j)}+u^0_{(j)}\right)=t^0_{(j)}\Big/n^0_{(j)}$. To ensure that the lowest dose is selected for the first cohort of subjects, $p^0_{(1)}$ is set equal to the TTL. The value chosen for $p^0_{(k)}$ is one which would be deemed too high if observed in the actual escalation procedure. This produces a prior dose-response curve that shows high doses to be unsafe. After fixing $p^0_{(1)}$ and $p^0_{(k)}$, the number of observations can be chosen to reflect the strength of the prior, and for each of the two dose levels is typically

chosen to be equal to the number of patients in a cohort. The number of DLTs $t^0_{(j)}$ can then be calculated as $n^0_{(j)} \times p^0_{(j)}$. Relative to the data collected during the study, the amount of pseudo data is small.

The posterior distribution also takes a Beta form since it is conjugate with the Binomial data. Both prior and posterior modal estimates can be obtained by fitting a logistic regression model using maximum likelihood methods in standard statistical software.

The initial allocation of doses depends directly on the prior information, usually fixed so that the lowest dose available has a mean probability of DLT at the TTL and the highest dose has some higher probability (in the case where the TTL is 0.2, the higher dose is often given a probability of 0.5). This ensures that the first cohort is administered the lowest possible dose. The prior information is implemented by the use of pseudo-data, where the number of toxicities and the number of observations are the parameters from the Beta distribution. Table 2-1 demonstrates this.

Dose, $d_{(j)}$	$n^0_{(j)}$	$t^0_{(j)}$	$p^0_{(j)} = t^0_{(j)} / n^0_{(j)}$
$d_{(1)}$	3	0.6	0.2
$d_{(k)}$	3	1.5	0.5

Table 2- 1: Pseudo-data for Logistic Regression Decision Procedure.

Logistic regression analysis is then conducted to obtain modal estimates (treating the prior/pseudo observations as true observations) and these estimates are used in the model to obtain modal estimates of the dose corresponding to the TTL. Different gain functions can then be used with these estimates to decide which dose to administer to the patients. In the initial step, the lowest dose is selected to administer to the first cohort for safety reasons regardless of whichever gain function is implemented. The

added benefit of having pseudo-data for the highest dose is that when its incorporated, the model is unlikely to choose high doses early in the escalation (where the weight of the pseudo-data has more influence), as there is already some evidence that high doses produce high probabilities of toxicity.

There are two commonly used gain functions, although there are more variations. The first is the patient gain [9, 13] (which is similar to the gain used in the CRM) which minimises the difference between the TTL and the probability of DLT for different doses found from the model. This is defined as:

$$g_{i(j)} = \left(\frac{1}{TTL - p_{i(j)}} \right)^2.$$

Where $p_{(j)}$ is the current estimated probability of DLT at dose $d_{(j)}$ after i patients. The dose $d_{(j)}$ that maximises this gain (produces the highest $g_{i(j)}$) is chosen to administer to the next cohort of patients.

When using the patient gain, all patients in a cohort are given the same dose since if one dose is believed to be best given the data so far, this is the same for all patients in the cohort.

Another example of a gain function is the variance gain. This involves finding the asymptotic subjective variance (i.e. the asymptotic variance calculated including the pseudo-data) of the estimate of the TD including the administrations of different possible doses for the next patient [14]. The next set of doses to be allocated is then based on the doses that will reduce the variance the most;

$$g_{i(J)} = \left(\frac{1}{\text{var}\left(\log\left(TD_{i,TTL}^{(+J)} \right) \right)} \right).$$

In this setting, TD_{TTL} is the dose that is believed to correspond exactly to the TTL and $TD_{i,TTL}^{(+J)}$ is the expected estimate of the TD_{TTL} after i patients, when incorporating the set of doses J for the next cohort of patients. The set of doses J can consist of different doses since a combination of doses may reduce the variance more than administering one dose to the entire cohort. The TD_{TTL} is expressed in terms of the parameters of the model and as the log transformed value since it is the log value of dose that is used in the model.

$$\log\left(TD_{TTL} \right) = \frac{\log\left(\frac{TTL}{1-TTL} \right) - \alpha}{\beta}. \tag{2.3}$$

The asymptotic variance for the estimate of the log(TD) is calculated using equation (2.3) and the likelihood function as in equation (2.1) of the CRM, where the function f is the logit link function as in equation (2.2). The likelihood for the calculation of the asymptotic variance includes extra expected observations from the set J of different possible doses to be administered. This is then defined as:

$$\text{var}\left(\log\left(TD_{i,TTL}^{(+J)} \right) \right) = \left[\left(\frac{\partial \log\left(TD_{TTL} \right)}{\partial \alpha}, \frac{\partial \log\left(TD_{TTL} \right)}{\partial \beta} \right) I_E^{-1} \begin{pmatrix} \frac{\partial \log\left(TD_{TTL} \right)}{\partial \alpha} \\ \frac{\partial \log\left(TD_{TTL} \right)}{\partial \beta} \end{pmatrix} \right],$$

where I_E^{-1} is the Expected Information Matrix found from twice differentiating the log-likelihood with respect to each parameter and taking the expectation of each element of the matrix. Only the elements involving the doses considered for administration in

the next cohort are affected by taking the expectation, since the expected values of $p_{(j)}$, $t_{(j)}$ and $n_{(j)}$ for those doses are included along with the observations from the data so far. On replacing the parameters with the estimated values after the i^{th} observation, the set of doses that produce the smallest variance of the estimated $\log(TD_{TTL})$ (increases the gain, $g_{i(j)}$) are allocated to the next cohort.

The patient gain is the more ethical gain function in the escalation procedure so this is what is usually used.

As with most dose-finding procedures, this method usually only incorporates information from the first cycle of therapy. As discussed with the CRM, in some therapeutic areas where late toxicities are common, this poses quite a problem. It would be unfeasible to wait until every patient has completed all cycles of therapy as this would result in an impractically long trial, however when there are many cycles of therapy it could be considered unethical to disregard all information after the first cycle of therapy and allocate doses based on just the first. This could result in allocating higher doses that would appear to be reasonably safe early in the treatment process but after an accumulation of dose could become dangerously toxic.

2.3 Overview

The traditional rule-based designs are extremely easy to follow, and clinicians understand the logic behind the rules. However they do not produce that much information about the actual dose-response relationship.

The CRM was the first of the model-based designs to be created and is therefore widely known and recognised. The two-stage CRM in particular allows cautious escalations until enough information is produced to safely move to the model-based escalation and can even be moved to a maximum likelihood model approach (CRML).

The extension of the CRM, the TITE-CRM, also allows inclusion of late occurring toxicities by way of weighting the likelihood contributions. The main issue with all of the CRM procedures (apart from the CRML) is that many complicated integrals need to be numerically calculated in order to obtain Bayes Estimates of the posterior mean. While this is still feasible through the use of MCMC approaches, it is not easily explainable or intuitive to non-statisticians, which may cause clinicians to opt for alternative, simpler approaches. The CRML does not have this issue but it is then questionable as to whether the slow start of this dose-escalation procedure is ethical, particularly when there are many possible dose-levels available. The converse of this is that by skipping the cautious escalation stage and using the one-stage CRM, the model is very unconservative, even when using pessimistic prior information. The one-stage CRM can easily allocate overly toxic doses early on in the procedure. The TITE-CRM would also need extensive exploratory investigation to find a weighting system that portrays the true effect of the drug over time.

The EWOC design aims to combat the CRM's issues of allocating overly toxic doses early in the procedure by making it more likely to allocate a dose below the true TD than above it. This is quite effective if the main issue is simply not to expose patients to overly toxic doses especially when used in conjunction with pessimistic prior information. A TD is estimated including the use of the pessimistic prior, and a dose even lower than this is most likely to be chosen to be administered to the patients. The issue with this procedure arises when considering the effect of administering sub-therapeutic doses. Not only is it unethical to administer sub-therapeutic doses to patients (which in cancer trials, the patients are cancer patients themselves) but if there are many dose-levels to consider, the trial could take a very long time. There is also an issue that (especially if the quantile is not chosen well) the estimated TD at the end of

the trial may actually be higher than any dose administered throughout the study. Suggesting a dose for further (and larger) studies, that has not actually been tested, seems unethical. Extensive knowledge about the dose-response relationship would be required to ensure this extrapolation is safe.

The LRDP seems the safest method discussed in this review. By incorporating pessimistic prior information about both the lowest and highest doses, the early allocations are unlikely to overdose patients, and once the first few observations are obtained, it is also unlikely to underdose by too much. Logistic regression analysis is also very simple to carry out as there are many software packages that carry out this type of analysis easily. The main issue with this procedure is one that is common to most of the other procedures also, that is the inclusion of observations from just the first cycle of therapy in the analysis. Although it would be unfeasible to wait for a patient to complete the whole treatment, when there are many cycles of therapy it seems wasteful to disregard so much information.

Chapter 3 explores some existing data from 38 Phase I dose-finding studies to provide motivation for the development of a new procedure which is then discussed in Chapter 4. A new model will be considered for the Bayesian Decision Procedure which allows the inclusion of later cycles of therapy and the derivation of the corresponding steps of the Bayesian decision procedure for this model will be presented.

3. Investigating data trends in Phase I trials

3.1 38 Phase I Trials in Molecularly Targeted Agents

This chapter investigates the trends in occurrences of DLTs across treatment cycles, and their association with different baseline characteristics such as age, gender and primary tumour type. This investigation is based on a dataset described in Postel-Vinay [1]. The dataset consists of 38 dose-finding trials that occurred at 2 different institutions, the Royal Marsden Hospital, UK and Institut Gustave-Roussy, France between January 2005 and July 2008. For all patients who participated in one of these trials, data were provided on age, gender, weight, height and tumour type. There was information on the treatment dose, the number of cycles of therapy each patient had received and for each occurrence of a toxic event, the grade of the event, the cycle in which it occurred, the family to which the toxicity belonged (renal etc.) and whether it was a DLT as defined by the protocol. Different drug types are investigated over these trials which have different mechanisms and routes of action, different schedules for administration, different numbers of dose levels and different MTDs.

Postel-Vinay et al. [1] investigate the rate of occurrence of toxicities for Molecularly Targeted Agents (MTAs). Many of the existing dose-finding procedures have been developed for the traditional chemotherapy where the MTD was deemed to be the most efficacious. The dose for investigation with MTAs may not be the maximum dose that is tolerated, but a biologically optimal dose which offers a similar efficacy profile to a higher dose but with a lower toxicity profile. The toxicity level of interest may then be redefined. Such a dose may then be administered for longer periods of time until the occurrence of disease progression or resistance, so the consideration of

toxicities occurring in later cycles of therapy may need to be given more attention than they are in current procedures.

Since the trials in the investigation are based on different drugs with differing MTDs and dose levels, some kind of normalisation of the doses is needed. Every patient that participated in a trial has data about the dose of drug that they received and the MTD of the drug for that trial. A normalised dose can be created by dividing each patient's dose by the MTD. For the 38 trials in the dataset, only 5 were on average administering the MTD to their patients (the average normalised dose was 1). 8 out of the 38 were administering on average a dose that was too high (the average normalised dose was >1) and 25 out of the 38 trials were administering doses that were lower than the MTD (the average normalised dose was <1). An average normalised dose of less than 1 is expected since escalation procedures all take some time to escalate to the required dose. Therefore many cohorts of patients would receive lower doses in the early stages of the procedure. Since the average normalized dose that is administered is expected to be less than 1, those trials that administer an average normalized dose higher than 1 imply that in fact they are either escalating to too high doses very quickly, or they are administering doses that are much higher than the lower doses with great frequency, such that the low doses required for early escalation are not as prevalent in the calculation of the average.

3.2 Data cleaning

Initial investigation of the datasets highlighted some interesting points. First, there were other grade 3 or 4 toxic events that occurred but were not included as DLTs as they were not toxic events as specified by the trial protocol. Second, all events that were listed as DLTs occurred in the first cycle of therapy. This is not unusual as this is generally how dose-escalation studies are conducted. However, there were also some

toxic events that occurred in cycle 1, and were part of the family of specified toxic events to be called a DLT, but were not listed as a DLT. Furthermore, there were some patients who experienced DLTs in a cycle who were then seemingly not removed from the trial as they then proceeded to contribute further toxicities, including grade 3-4 toxicities, in later cycles also. While it is expected to continue to treat patients after their first DLT, although likely at a reduced dose, and continue to observe said patients for toxicities, this is for ethical and safety reasons rather than for use in the analysis. Therefore subsequent toxicities cannot be classed as DLTs. There were also patients that contributed more than 1 grade 3 or 4 toxic event to one cycle. Since the focus of this investigation is to consider a patient's first occurrence of a DLT, some data cleaning needs to be conducted in order to obtain the data in a format relevant to this.

A new binary variable was created to include any grade 3 or 4 toxic event in any cycle of therapy as an event of interest. Since each trial had different drugs and doses, it can be assumed that different toxicities as specified in the protocol were recorded as DLTs. Therefore, looking at all grade 3 or 4 toxicities will give more of an overview as to the prevalence of severe toxicities. Furthermore, DLTs only occurred in cycle 1, so by looking at all grade 3 or 4 toxicities some insight can be gained into how toxicities occur over cycles. There were also some events that were listed as a DLT despite a grade 3 or 4 toxicity not occurring. In practice, DLTs are not necessarily just the occurrence of one grade 3-4 toxicity, but may be defined as multiple or recurrent incidences of lower grade toxicities. These have therefore been left as DLTs. This new variable therefore comprises all protocol specified DLTs and all other grade 3 or 4 toxicities in all cycles. For this investigation, the new variable will be called DLT^{+} and

will be assigned a value of 1 to all occurrences of a DLT^+ and 0 for all other observations.

Once the DLT^+ were defined, the patients who contributed more than 1 grade 3 or 4 toxic event in multiple cycles had all but the first event removed as they should not have contributed to the analysis after experiencing an event of that grade. In order to make the overall dataset consistent with this, any cycles before the first occurrence of a DLT^+ were assigned a non-event ($DLT^+=0$) and any cycles after the first occurrence of a DLT^+ were removed.

3.3 Occurrence of toxicities

Table 3-1 shows the number of patients who started each cycle along with the number of patients experiencing their first DLT^+ in that cycle, and the corresponding percentages of the number of patients experiencing their first DLT^+ in that cycle compared to the number of patients starting each cycle. The 7-10 category is the number of patients who started cycle 7 and likewise for 11-20. Any cycles after the 20^{th} cycle were omitted since there were very few patients in these later cycles and it was unclear as to the time period that the events actually occurred in (anything from cycle 20 up to cycle 33).

Cycle	#Patients	% of total patients, n=445	#Patients with first DLT^+	% of Patients entering cycle
1	445	100%	38	8.5%
2	331	74.4%	14	4.2%
3	177	39.8%	8	4.5%
4	118	26.5%	3	2.5%
5	66	14.8%	3	4.5%
6	49	11.0%	1	2.0%
7-10	37	8.3%	0	-
11-20	15	3.4%	1	0.67%*

Table 3- 1: Patients experiencing their first DLT^+ in each cycle. *average percentage per cycle for 10 cycles.

As can be seen, the number of patients who start each cycle of treatment decreases with every cycle, with the amount of reduction between cycles generally decreasing throughout the trial. More events tend to occur in early cycles so more patients do not proceed to the next cycle, whereas fewer events occur later so fewer patients need to stop treatment. The exception here is for patients proceeding from cycle 2 to cycle 3. The reduction in patients here is the largest with just over half of the patients entering cycle 2 proceeding to cycle 3. The proportion of patients who experience their first event in each cycle tends to decrease with cycle. When looking at the later cycles that have been grouped together, an average percentage per cycle is displayed as the percentage of patients experiencing $DLTs^{+}$ divided by the number of cycles in the grouped category. The number of protocol specified DLTs that occur in cycle 1 is 26, which is just 5.8% of the patients starting cycle 1.

It can be seen that 60/68 first occurrences of $DLTs^{+}$ occur in the first 3 cycles compared to 38/68 occurring in cycle 1. The proportion of first $DLTs^{+}$ occurring in cycle 2 is approximately half of that in cycle 1. Later cycles also seem to have smaller proportions too suggesting the proportion of patients having first $DLTs^{+}$ in each cycle decreases as patients last longer in a trial.

Of the 38 trials included in the dataset, some had very few observations of $DLTs^{+}$ and often only had one type of response (no events). Most of the trials contributed some events and non-events but there was only one trial (Trial 64) that contributed a good amount of information. This trial is therefore looked into more closely.

Of the 35 patients in trial 64, 17 patients had events with 11 patients having protocol specified events and there were 14 DLTs in cycle 1. The trial had 121 patient cycles (121 cycles of therapy across the 17 patients) of information and 52 $DLTs^{+}$ occurred

in total. Since DLTs$^+$ were defined as the first Grade 3 or 4 toxicity, this implies that several patients had more than one in the same cycle, likely of a different type. Focus therefore lies on the patients who experience a DLT$^+$ rather than the number itself. The breakdown of the trial and the events occurring throughout is shown in Table 3-2.

Cycle	#Patients	#Patients with first DLT$^+$	Patients with DLT$^+$ as % of Patients entering cycle
1	35	7	20.0%
2	26	4	15.4%
3	10	3	33.3%
4	6	1	16.7%
5	5	2	40.0%
6	3	-	-
7-10	1	-	-
11-20	0	-	-

Table 3- 2: Trial 64 – Patients experiencing their first DLT$^+$ in each cycle.

As can be seen, over half of the patients who have a DLT$^+$ experienced it after cycle 1. None occur in the very late cycles so these could be disregarded but over 80% occur within the first 3 cycles. The proportion of patients experiencing their first DLT$^+$ is reasonably stable for the first 4 cycles with a sudden surge in the 5th.

Trial 64 shows similar trends to those shown in Table 3-1, however due to the small sample size a precise trend cannot be determined. Therefore the results from Table 3-1 will be used to aid the choice of model parameter values in the simulations described in chapters 5, 6, 7 and 8. To generalise, the probability of a DLT occurring for patients who proceed to later cycles will be set to be half of the probability of a DLT in the previous cycle. The first 3 cycles of therapy will be considered.

3.4 Covariate EDA

Investigation of some of the patient characteristics could provide insight into how a personalised dose allocation procedure could be used to give the dose that is best for each individual patient.

This part of the EDA will focus on the first occurrence of $DLTs^{+}$ in different cycles of therapy for different levels of a factor representing patient groups. This will provide a general idea of the prevalence of toxic events for all cycles for different levels or categories of a covariate and will provide a more specific understanding of the pattern of toxic events across cycles for different levels or categories.

3.4.1 Age

The mean ages of all patients entering each cycle and also just patients who experienced their first DLT^{+} during the cycle are shown in Table 3-3.

Cycle, n	Mean Age	S.D	Min	Max
Cycle 1, n=445	56.10	12.49	18	86
n^{DLT+}=38	55.36	13.79	29	78
Cycle 2, n=331	56.11	12.21	18	86
n^{DLT+}=14	57.79	8.31	40	71
Cycle 3, n=177	56.95	12.60	20	86
n^{DLT+}=8	50.13	13.11	32	68
Cycle 4, n=118	56.90	11.90	20	79
n^{DLT+}=3	62.67	5.51	59	69
Cycle 5, n=66	57.61	12.77	25	79
n^{DLT+}=3	60.00	4.58	56	65
Cycle 6, n=49	57.71	12.77	25	79
n^{DLT+}=1	60.00	-	-	-
Cycle 7-10, n=37	57.95	12.73	25	79
n^{DLT+}=0	-	-	-	-
Cycle 11-20, n=15	55.13	12.56	27	73
n^{DLT+}=1	62	-	-	-

Table 3- 3: Mean age of patients in each cycle for all patients, and patients with DLT^{+}.

The mean age for all patients is quite consistent across cycles. For patients experiencing their first DLT^{+}, the mean age is generally younger in the earlier cycles

and slightly older in the later cycles. If focussing on cycles 1-3, it could be generalised that younger patients tend to have DLTs^{+}.

A further way to investigate the effect of age on the occurrence of DLT^{+} is to look at the occurrences in various age categories. In the dataset, the following four age categories had been created <50, 51-58, 59-65, >65.

Table 3-4 shows the number of patients in each age category.

Age Category	**<50**	**51-58**	**59-65**	**>65**
#Patients entering study	131	95	125	94

Table 3- 4: Number of patients in each age category.

The number of patients in each age category differs slightly, but there are reasonably equal proportions in the first 2 categories compared to the last 2 categories.

Table 3-5 shows the number of patients in each age category starting each cycle, the number of patients experiencing DLTs^{+} in each cycle which is also displayed as a proportion of the total patients starting each cycle.

Cycle, n	**Age category**	**<50**	**51-58**	**59-65**	**>65**
Cycle 1, n=445 n^{DLT+}=38	#Patients #Pats with 1st DLT+, (% of pats in cycle)	131 14 (10.7%)	95 5 (5.3%)	125 9 (7.2%)	94 10 (8.8%)
Cycle 2, n=331 n^{DLT+}=14	#Patients #Pats with 1st DLT+, (% of pats in cycle)	94 3 (3.2%)	76 3 (3.9%)	96 7 (7.3%)	65 1 (1.5%)
Cycle 3, n=177 n^{DLT+}=8	#Patients #Pats with 1st DLT+, (% of pats in cycle)	47 4 (8.5%)	35 1 (2.9%)	54 2 (3.7%)	41 1 (2.4%)
Cycle 4, n=118 n^{DLT+}=3	#Patients #Pats with 1st DLT+, (% of pats in cycle)	29 0 (0.0%)	28 0 (0.0%)	36 2 (5.6%)	25 1 (4.0%)

Table 3- 5: Numbers of patients with their first DLT^{+} in each cycle for different age categories. *average percentage per cycle for 10 cycles.

Cycle 5, n=66 $n^{DLT+}=3$	#Patients #Pats with 1st DLT+, (% of pats in cycle)	16 0 (0.0%)	15 1 (6.7%)	21 2 (9.5%)	14 0 (0.0%)
Cycle 6, n=49 $n^{DLT+}=1$	#Patients #Pats with 1st DLT+, (% of pats in cycle)	12 0 (0.0%)	8 0 (0.0%)	16 1 (6.3%)	13 0 (0.0%)
Cycle 7-10, n=37 $n^{DLT+}=0$	#Patients #Pats with 1st DLT+, (% of pats in cycle)	9 0 (0.0%)	5 0 (0.0%)	13 0 (0.0%)	10 0 (0.0%)
Cycle 11-20, n=15 $n^{DLT+}=1$	#Patients #Pats with 1st DLT+, (% of pats in cycle)	6 0 (0.0%)	1 0 (0.0%)	5 1 (2.0%*)	3 0 (0.0%)

Table 3-5 cont.: Numbers of patients with their first DLT^+ in each cycle for different age categories. *average percentage per cycle for 10 cycles.

Quite a high proportion of patients in the youngest age group experience their first DLT^+ during the first 3 cycles, particularly compared to the oldest group. The youngest group have no events after cycle 3, whereas the older groups continue to have small numbers of events into later cycles. Generally the proportion of patients having their first DLT^+ decreases over cycles for all age categories.

Based on the results from section 3.3, the investigation of the dose escalation procedures can be focused on the first 3 cycles of therapy. Therefore, the generalised pattern adopted from the investigation of age for use in this thesis is that younger patients have a higher chance of DLT than older patients, and the occurrence of DLTs decreases equally over cycles for the different age groups.

3.4.2 Gender

Comparing the occurrence of DLT^+ for males and females regardless of when the toxic event occurred gives a general idea of the overall prevalence for each gender. The frequency table in Table 3-6 shows the number of male and female patients along with the number and proportion of those who had any level of toxic event.

Gender	Female	Male
# Patients	190	255
# Pats with first DLT+ (% of total)	35 (18.4%)	33 (12.9%)
# Pats with first protocol DLT in cycle 1 (% of total)	14 (7.4%)	16 (6.3%)

Table 3- 6: Number of patients with a DLT^+ and a protocol specified DLT for each gender.

There are a slightly larger number of males included in the trials, but a smaller proportion of those experience $DLTs^+$ when compared to females. When considering just the protocol specified DLTs in cycle 1, there are notably fewer events than when considering $DLTs^+$ and they are more evenly spread across the genders.

Further investigation can be conducted to see if the time of occurrence of the first DLT^+ is also different for males and females.

Cycle, n	Gender	Females	Males
Cycle 1, n=445 n^{DLT+}=38	#Patients	190	255
	#Pats with 1st DLT+, (% of pats in cycle)	20 (10.5%)	18 (7.1%)
Cycle 2, n=331 n^{DLT+}=14	#Patients	137	194
	#Pats with 1st DLT+, (% of pats in cycle)	7 (5.1%)	7 (3.6%)
Cycle 3, n=177 n^{DLT+}=8	#Patients	71	106
	#Pats with 1st DLT+, (% of pats in cycle)	4 (5.6%)	4 (3.8%)
Cycle 4, n=118 n^{DLT+}=3	#Patients	50	68
	#Pats with 1st DLT+, (% of pats in cycle)	1 (2.0%)	2 (2.9%)
Cycle 5, n=66 n^{DLT+}=3	#Patients	34	32
	#Pats with 1st DLT+, (% of pats in cycle)	1 (2.9%)	2 (6.3%)
Cycle 6, n=49 n^{DLT+}=1	#Patients	26	23
	#Pats with 1st DLT+, (% of pats in cycle)	1 (3.8%)	0 (0.0%)

Table 3- 7: Number of patients with first DLT^+ in each cycle for each gender. *average percentage per cycle for 10 cycles.

Cycle 7-10, n=37 $n^{DLT+}=0$	#Patients	19	18
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	0 (0.0%)
Cycle 11-20, n=15 $n^{DLT+}=1$	#Patients	7	8
	#Pats with 1st DLT+, (% of pats in cycle)	1 (1.4%*)	0 (0.0%)

Table 3-7 cont.: Number of patients with first DLT^+ in each cycle for each gender. *average percentage per cycle for 10 cycles.

Both males and females have generally a decreasing number of $DLTs^+$ over cycles. For the first 3 cycles, the proportion of patients in each cycle having $DLTs^+$ is higher for females. Cycles 4 and 5 have a slightly higher proportion of male patients having $DLTs^+$ but after cycle 5 males have no more $DLTs^+$ whereas females continue to have DLTs until very late cycles of therapy.

Despite having a smaller number of female patients overall, more females are experiencing events for much longer. Both genders have a steadily decreasing number of $DLTs^+$ over cycles.

By focusing again on the first 3 cycles of therapy, the results can be summarised as females have a higher chance of DLT than males with both groups having an equal reduction of DLTs with successive cycles.

3.4.3 Primary Tumours

The differences in occurrences of $DLTs^+$ between age groups and genders could be due to the primary tumour that is being treated. As an example, treatments associated with breast or gynaecological tumours may cause more toxicities and are observed only in women. Furthermore, tumours occurring in older patients may not result in as many toxicities, perhaps due to a higher tolerance.

The occurrence of toxicities according to primary tumours can be investigated to try to aid understanding of the differences between age groups and gender but in reality would not be accounted for in the analysis of toxicities. Usually a specific trial would

treat one certain type of tumour so differences between patients would not be due to tumour type. The understanding of tumour type toxicities for different groups of patients can be used to aid the design of the trial, deciding how long to observe patients for and perhaps what kind of toxicity to expect.

Table 3-8 shows the occurrence of toxicity for different primary tumours.

Tumour Type	Number of patients	No. patients with DLT+ (% of total pats)	No. patients with DLTs (% of total pats)
Prostate	61	2 (3.3%)	1 (1.6%)
Sarcoma	49	6 (12.2%)	4 (8.2%)
Breast & Gynaeco. (B&G)	40	5 (12.5%)	3 (7.5%)
Central Nervous System (CNS)	7	-	-
Urological	9	-	-
Gastrointestinal (GI)	151	11 (7.3%)	7 (4.6%)
Thoracic	77	12 (15.6%)	9 (11.7%)
Melanoma	23	2 (8.7%)	2 (8.7%)
Other	28	-	-

Table 3- 8: Numbers of patients, and number with DLT^+ and protocol specified DLTs for each primary tumour type.

Firstly, since the sample sizes for the CNS and Urological tumour types are very small and neither produce DLTs or $DLTs^+$, these can be hereafter excluded from further exploratory data analysis (EDA) as no information will be gained from them. The 'Other' group will also be excluded since there is no information about the primary tumour and it can be assumed that multiple 'other' tumour types have been grouped together due to small numbers of patients.

The largest number of patients have a GI primary tumour, however quite a modest proportion of these patients actually have DLTs or $DLTs^+$. The highest proportion of patients experiencing DLTs and $DLTs^+$ occur for patients with thoracic tumours and the lowest proportion occurs for patients with prostate tumours. Breast and

Gynaecological, and Sarcoma primary tumours also have a slightly larger proportion of patients having DLTs.

The pattern in which the DLTs occur for each primary tumour type is shown in Table 3-9.

Cycle, n	Primary	Prostate	Sarcoma	B&G
Cycle 1, n=445 n^{DLT+}=38	#Patients #Pats with 1st DLT+, (% of pats in cycle)	61 1 (1.6%)	49 4 (8.2%)	40 3 (7.5%)
Cycle 2, n=331 n^{DLT+}=14	#Patients #Pats with 1st DLT+, (% of pats in cycle)	48 0 (0.0%)	38 1 (2.6%)	30 4 (13.3%)
Cycle 3, n=177 n^{DLT+}=8	#Patients #Pats with 1st DLT+, (% of pats in cycle)	27 1 (3.7%)	22 1 (4.5%)	12 0 (0.0%)
Cycle 4, n=118 n^{DLT+}=3	#Patients #Pats with 1st DLT+, (% of pats in cycle)	20 1 (5.0%)	18 1 (5.6%)	10 0 (0.0%)
Cycle 5, n=66 n^{DLT+}=3	#Patients #Pats with 1st DLT+, (% of pats in cycle)	10 0 (0.0%)	10 0 (0.0%)	6 0 (0.0%)
Cycle 6, n=49 n^{DLT+}=1	#Patients #Pats with 1st DLT+, (% of pats in cycle)	5 0 (0.0%)	7 0 (0.0%)	5 0 (0.0%)
Cycle 7-10, n=37 n^{DLT+}=0	#Patients #Pats with 1st DLT+, (% of pats in cycle)	5 0 (0.0%)	4 0 (0.0%)	4 0 (0.0%)
Cycle 11-20, n=15 n^{DLT+}=1	#Patients #Pats with 1st DLT+, (% of pats in cycle)	4 0 (0.0%)	1 0 (0.0%)	1 0 (0.0%)

Table 3- 9: Number of patients with first DLT^+ in each cycle of occurrence for each primary tumour type. *average percentage per cycle for 10 cycles.

		Melanoma	GI	Thoracic
Cycle 1, n=445 n^{DLT+}=38	#Patients	23	151	77
	#Pats with 1st DLT+, (% of pats in cycle)	2 (8.7%)	7 (4.6%)	9 (11.7%)
Cycle 2, n=331 n^{DLT+}=14	#Patients	17	110	57
	#Pats with 1st DLT+, (% of pats in cycle)	1 (5.9%)	5 (4.5%)	3 (5.3%)
Cycle 3, n=177 n^{DLT+}=8	#Patients	9	57	32
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	3 (5.3%)	3 (9.4%)
Cycle 4, n=118 n^{DLT+}=3	#Patients	6	34	20
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	0 (0.0%)	1 (5.0%)
Cycle 5, n=66 n^{DLT+}=3	#Patients	4	15	15
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	0 (0.0%)	3 (20.0%)
Cycle 6, n=49 n^{DLT+}=1	#Patients	2	14	12
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	0 (0.0%)	1 (8.3%)
Cycle 7-10, n=37 n^{DLT+}=0	#Patients	2	9	10
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Cycle 11-20, n=15 n^{DLT+}=1	#Patients	0	2	5
	#Pats with 1st DLT+, (% of pats in cycle)	0 (0.0%)	0 (0.0%)	1 (2.0%*)

Table 3- 9 cont.: Number of patients with first DLT^+ in each cycle of occurrence for each primary tumour type. *average percentage per cycle for 10 cycles.

By splitting up by tumour type, it actually seems that the occurrence of $DLTs^+$ generally decrease after cycle 1, but then occur at a slightly increasing rate over subsequent cycles. Very few of the tumour types have patients with $DLTs^+$ in the very late cycles, in fact only thoracic tumours have any after cycle 4.

Thoracic and Sarcoma tumour types have patients with a higher proportion of $DLTs^+$ occurring in cycle 1-3 and some of the tumours only have patients with $DLTs^+$ for the first 2 cycles. It should therefore be concluded that dependent on the tumour type investigated in the trial, the length of the observation period should be adapted along with the TTL.

3.4.4 Interaction between Gender/Age and Primary Tumour Type

The EDA carried out for the gender covariate prompted the question as to whether the occurrence of later DLTs for females could be to do with the type of tumour that was being treated. However, EDA for the primary tumour type shows this not to be the case. The only tumour type that is specific only to women is the breast or gynaecological tumour, and these only experienced DLTs in the first 2 cycles. The only tumour type in which patients experienced DLTs in the later cycles (as the females did) was the thoracic tumour, which is not specific to women. Furthermore, looking at tumour types specific to men such as the prostate tumour type does not provide any insight either since there are so few DLTs occurring for prostate cancer, and these only occur up until cycle 4. Therefore, there is not enough information to conclude that tumour type does has a confounding effect on the effect caused by gender to the time of the DLT.

When considering age also, there was very little difference between tumour type and age over cycles.

As discussed, in reality different tumour types would not be included in the same trial so would not be adjusted for within the analysis and escalation procedure. They can however be used in the decision process for designing the trial in order to consider what kind of toxicities would occur and when. This could aid the decision on the observation period for toxicities, and what TTL should be considered.

3.4.5 Family of Toxicity

The timing and seriousness of a DLT in a particular study may depend on the type (family) of toxicity being assessed. For example, certain types of toxicity (e.g. gastrointestinal: vomiting, diarrhoea etc.) may be more likely to occur particularly in early or late cycles, whereas another type (e.g. cutaneous: rash etc.) may be different,

and furthermore, there may be a relationship between timing and seriousness, e.g. certain types of toxicity may be more toxic early on.

For this investigation, the occurrence of a DLT^+ is not just a yes or no outcome, the actual family of toxicity is of interest. Therefore the first occurrence of each family of DLT^+ is included. The sum of the number of patients with each family of toxicity may therefore be greater than the total number of patients experiencing any type of DLT^+. Table 3-10 shows when these events occur.

Cycle, n	Family	Cutaneous	GI	Renal
Cycle 1,	#Pats with 1st DLT+,	13	31	7
n=445	% of n	2.9%	7.0%	1.6%
n^{DLT+}=38	% of n^{DLT+}	34.2%	81.6%	18.4%
Cycle 2,	#Pats with 1st DLT+,	3	14	1
n=331	% of n	0.9%	4.2%	0.3%
n^{DLT+}=14	% of n^{DLT+}	21.4%	100%	7.1%
Cycle 3,	#Pats with 1st DLT+,	3	4	3
n=177	% of n	1.7%	2.3%	1.7%
n^{DLT+}=8	% of n^{DLT+}	37.5%	50.0%	37.5%
Cycle 4,	#Pats with 1st DLT+,	0	2	1
n=118	% of n	-	1.7%	0.8%
n^{DLT+}=3	% of n^{DLT+}	-	66.7%	33.3%
Cycle 5,	#Pats with 1st DLT+,	1	3	2
n=66	% of n	1.5%	4.5%	3%
n^{DLT+}=3	% of n^{DLT+}	33.3%	100%	66.7%
Cycle 6,	#Pats with 1st DLT+,	0	0	1
n=49	% of n	-	-	2.0%
n^{DLT+}=1	% of n^{DLT+}	-	-	100%
Cycle 7-10,	#Pats with 1st DLT+,	0	0	0
n=37	% of n	-	-	-
n^{DLT+}=0	% of n^{DLT+}	-	-	-
Cycle 11-20,	#Pats with 1st DLT+,	0	1	0
n=15	% of n	-	-	-
n^{DLT+}=1	% of n^{DLT+}	-	100%	-

Table 3- 10: Number of patients with a first DLT^+ in each cycle for each type of toxicity.

The family of toxicity that has the highest level of occurrence is the GI family and this is true for every cycle. The proportion of patients experiencing each type of toxicity in each cycle is generally quite consistent across cycles for cutaneous and renal toxicities

and decreases with cycle for GI toxicities. Of the patients who do experience a DLT^+ it is most likely to be a GI toxicity and the chance of a DLT^+ being GI or cutaneous is quite consistent across cycles. Although the chance of a DLT^+ being a renal toxicity is generally very low, it does tend to increase with cycles and a higher proportion of patients experiencing $DLTs^+$ experience renal $DLTs^+$ in later cycles than in earlier cycles.

As for primary tumour types, it is not reasonable to account for the type of toxicity occurring in the analysis since an exhaustive list of toxicities would not be feasible to adjust for, however, in accounting for which type is most likely and considering the severity and when they might occur can again aid the decision process of designing the trial and analysis plan.

4. Methodology – ICSDP

From the evidence produced in Chapter 3, a new model is proposed to incorporate data from later cycles of therapy within a Bayesian decision framework. This model is the Interval-Censored Survival (ICS) Model. This Chapter therefore presents a new Bayesian decision procedure which incorporates the use of the ICS model. This procedure is called the Interval-Censored Survival Decision Procedure (ICSDP).

4.1 The Interval-Censored Survival Model

The ICS Model is derived from the proportional hazards assumption as shown in Collett [15]. It has traditionally been used with analysing time-to-event outcomes in order to reduce bias associated with the uncertainty of the exact timings of events. For example, in the analysis of progression free survival (PFS), the event of disease progression is usually determined at scheduled visits and assessments. The exact time of progression is therefore unknown and assigned to the next scheduled assessment. When symptoms associated with progression arise, usually for control treatments, unscheduled visits and assessments occur which implies that progression can be detected earlier and the PFS is therefore obtained accurately. If a control treatment estimates the PFS accurately, the hazard associated with the control treatment is also estimated accurately. The occurrence of symptoms associated with progression may be suppressed for the experimental treatment and will therefore not be observed. The occurrence of progression may then not be observed until the scheduled assessment, which overestimates the PFS, which in turn, underestimates the hazard associated with the experimental treatment. The overall estimated hazard ratio (HR) will then be underestimated and the experimental treatment may look to perform better than it actually does. By allowing inclusion of observations at each interval, the uncertainty of the exact time of progression is localised to one specific period. This therefore

reduces the bias observed when unscheduled assessments occur intermittently with scheduled assessments.

The idea of observing events occurring in different intervals can be directly related to the design of Phase I dose-finding trials. Traditionally, dose-finding studies observe the occurrence of a patient's first DLT during one fixed period of time. Patients should remain on therapy after the first cycle of treatment, but only observations from the first cycle will be used for analysis. The observations are binary, either they had a DLT in cycle 1 or they did not, and it is therefore very easy to analyse these events over a fixed period of time, either with a model-based analysis or rule-based. The first cycle of therapy is also the cycle for which the first DLT for a patient is expected to occur with highest frequency, although they can be expected to occur at a decreasing rate of frequency over time. By only using one cycle of therapy for the analysis, the trials are often very short in time. To observe for multiple cycles of therapy would increase the duration of the trial and when observing binary responses, the same issue arises as described for analysing PFS. The time of DLT would not necessarily be captured if multiple cycles were combined to one fixed period of time, and the probability of DLT (P(DLT)) would be assumed to be constant for the whole time period. This may underestimate a patient's chance of experiencing a DLT early in the treatment phase and overestimate it later on.

The ICS model now becomes attractive since it can look at a larger fixed period of time, and break it down into intervals, in this case cycles. The binary endpoint of whether a patient's first DLT occurs or not is still utilised, however the endpoint is now whether a DLT occurred in a given cycle, and the occurrence of DLTs will now be dependent on the cycles that occur previously and a DLT not occurring. A patient will contribute information to the analysis for all cycles of therapy they complete, up

to and including the cycle during which they have a DLT. By allowing patients to contribute information for every completed cycle of therapy, this model allows for non-informative withdrawal by including cycles up to the first DLT or withdrawal. Although dropouts before the occurrence of a DLT are not particularly expected in this phase of development, it is still an issue that should be considered since cancer patients may experience progression so therefore may be withdrawn, and since the sample size for Phase I trials is so small already.

For Phase I trials, interest lies in modelling the probability of a patient having their first DLT, on dose level $d_{(j)}, j = 1,...,k$ denoted by $p_{(j)}$. Traditionally, this is the probability of having a DLT during the first cycle of therapy. Interest may lie in assessing the probability of a DLT over s cycles of therapy, where each cycle of therapy $l, l = 1,...,s$, with s being the maximum number of cycles, begins at time c_{l-1} and finishes at time c_l with $c_0 = 0$. Let $p_{(j)l}$ be the probability of a patient experiencing their first DLT on dose level $d_{(j)}$ during cycle l and $p_{(j)s}{}^{c}$ be the probability that no DLT occurs during the first s cycles, i.e. the complement of $p_{(j)s}$. The cumulative probability of an event occurring during the first l cycles for a patient on dose level $d_{(j)}$ can then be defined as:

$$p_{(j)}(c_l) = \sum_{m=1}^{l} p_{(j)m} .$$

Here, $p_{(j)m}$ for $m = 1,...,l$ are probabilities relating to mutually exclusive events. Therefore the sum from $m = 1,...,s+1$ is equal to 1.

The ICS model is based on the probability of a DLT occurring during a given cycle conditional on the fact that there has been no DLT in previous cycles. The conditional

probability of having an event in cycle l (after time c_{l-1}) given that there has been no DLT in any interval prior to cycle l is defined as $\pi_{(j)l} = P\left(c_{l-1} < T_{(j)} \le c_l \mid T_{(j)} > c_{l-1}\right)$ where $T_{(j)}$ is the true time at which a DLT occurs on dose level $d_{(j)}$. The conditional probabilities can then be combined to calculate the unconditional probabilities of DLT $p_{(j)l}$ for a given cycle l and dose level $d_{(j)}$ as follows:

$$\begin{aligned}
p_{(j)l} &= P\left(c_{l-1} < T_{(j)} \le c_l\right) \\
&= P\left(c_{l-1} < T_{(j)} \le c_l \mid T_{(j)} > c_{l-1}\right) P\left(T_{(j)} > c_{l-1}\right) \\
&= \pi_{(j)l} P\left(T_{(j)} > c_{l-1}\right) \\
&= \pi_{(j)l} P\left(T_{(j)} > c_{l-1} \mid T_{(j)} > c_{l-2}\right) P\left(T_{(j)} > c_{l-2}\right) \\
&= \pi_{(j)l}\left[1 - P\left(c_{l-2} < T_{(j)} \le c_{l-1} \mid T_{(j)} > c_{l-2}\right)\right] P\left(T_{(j)} > c_{l-2}\right) \\
&= \pi_{(j)l}\left(1 - \pi_{(j),l-1}\right) P\left(T_{(j)} > c_{l-2}\right) \\
&\ \ etc.
\end{aligned}$$

This can then be generalised to the following:

$$p_{(j)l} = \begin{cases} \pi_{(j)l} & l = 1 \\ (1-\pi_{(j)1})(1-\pi_{(j)2})\ldots(1-\pi_{(j),l-1})\pi_{(j)l} & l = 2, \ldots, s \\ (1-\pi_{(j)1})(1-\pi_{(j)2})\ldots(1-\pi_{(j),l-2})(1-\pi_{(j),l-1}) & l > s. \end{cases}$$

It can be noted that $\pi_{(j),s+1} = 1$ since it is assumed that if the patient remained on treatment after surviving the first s cycles, at some point in the interval (c_s, ∞) a DLT would occur. This then explains why the unconditional probability of an event occurring in the interval (c_s, ∞) reduces as above. This arrangement of probabilities means that the likelihood can be constructed in terms of the conditional probabilities.

$$\prod_{j=1}^{k}\prod_{l=1}^{s+1} p_{(j)l}^{\,t_{(j)l}} = \prod_{j=1}^{k}\prod_{l=1}^{s+1}\left[\left(1-\pi_{(j),l-(l-1)}\right)\left(1-\pi_{(j),l-(l-2)}\right)\ldots\left(1-\pi_{(j),l-1}\right)\pi_{(j)l}\right]^{t_{(j)l}}$$

$$= \prod_{j=1}^{k}\pi_{(j),s+1}^{\,t_{(j),s+1}}\prod_{l=1}^{s}\pi_{(j)l}^{\,t_{(j)l}}\left(1-\pi_{(j)l}\right)^{q_{(j)l}} \qquad (4.1)$$

$$= \prod_{j=1}^{k}\prod_{l=1}^{s}\pi_{(j)l}^{\,t_{(j)l}}\left(1-\pi_{(j)l}\right)^{q_{(j)l}}$$

Here $t_{(j)l}$ is equal to the number of toxicities observed on dose level $d_{(j)}$ during cycle l and $q_{(j)l}$ is equal to the number of patients who have completed l cycles of therapy without experiencing a DLT $\left(n_{(j)l}-t_{(j)l}\right)$.

This is a Binomial likelihood so a generalised linear model can be used to model these probabilities. The link function for this generalised linear model can be defined from the proportional hazards assumption via the following mechanism (as seen in Collett [15]).

Redefining the conditional probabilities in terms of survival probabilities is shown below;

$$\pi_{(j)l} = P\left(c_{l-1} < T_{(j)} \le c_l \mid T_{(j)} > c_{l-1}\right)$$
$$= \frac{S_{(j)}(c_{l-1}) - S_{(j)}(c_l)}{S_{(j)}(c_{l-1})},$$

Where $S_{(j)}(c_l)$ is the survival probability (i.e. the probability of 'surviving' the cycle without experiencing an event) associated with dose level (j) by the end of cycle l.

This can be simplified to:

$$1-\pi_{(j)l} = \frac{S_{(j)}(c_l)}{S_{(j)}(c_{l-1})}.$$

The proportional hazards assumptions is defined as;

$$S_{(j)}(c_l) = \left[S_0(c_l)\right]^{e^{\eta_{(j)}}},$$

Where $\eta_{(j)}$ is the linear predictor of covariates associated with dose level $d_{(j)}$ such as $\eta_{(j)} = \theta \log(d_{(j)})$, and $S_0(c_l)$ is the baseline survival function at the end of cycle l, i.e. the survival probability associated with dose level $d_{(0)}$ at the end of cycle l. When the dose is transformed, this transformation of this dose level $d_{(0)}$ will be equal to 0, i.e. to apply a log transformation to $d_{(0)}$, $d_{(0)}$ will be equal to 1 such that once transformed is equal to 0. Applying the proportional hazards assumption to the above rearrangement gives;

$$
\begin{aligned}
1 - \pi_{(j)l} &= \left[\frac{S_0(c_l)}{S_0(c_{l-1})} \right]^{e^{\eta_{(j)}}}, \\
\log(-\log(1 - \pi_{(j)l})) &= \eta_{(j)} + \log\left(-\log\left(\frac{S_0(c_l)}{S_0(c_{l-1})} \right) \right) \\
&= \eta_{(j)} + \gamma_l
\end{aligned}
\tag{4.2}
$$

This link function is a complementary log-log link function, which includes a term that is dependent solely on the interval during which the event occurred. This is a factor with s levels which therefore allows separate intercept terms to be estimated for each cycle and therefore can allow for a differing dose-response relationship with time. The intercepts and the log(dose)-coefficient will all be log-hazard ratios comparing dose level $d_{(j)}$ to $d_{(0)}$.

Interest lies in estimating the dose that corresponds to a pre-defined probability of toxicity after s cycles of therapy. This can be defined as;

$$
\begin{aligned}
&p_{(j)}(c_s) \\
&= \begin{cases} \pi_{(j)1} & s = 1 \\ \pi_{(j)1} + \{(1 - \pi_{(j)1})\pi_{(j)2}\} + \cdots + (1 - \pi_{(j)1}) \ldots (1 - \pi_{(j),s-1})\pi_{(j)s} & s > 1 \end{cases}
\end{aligned}
$$

By putting in a rearrangement of the link function (in terms of $\pi_{(j)s}$) such that $\eta_{(j)} = \theta \log(d_{(j)})$, $p_{(j)}(c_s)$ in terms of the parameters can be found to be;

$$p_{(j)}(c_s) = 1 - \exp\left\{ d_{(j)}{}^{\theta} \left[-e^{\gamma_1} - e^{\gamma_2} - \ldots - e^{\gamma_s} \right] \right\},$$

which can then be rearranged in terms of the dose;

$$d_{(j)} = \exp\left\{ \frac{\log\left[\frac{\log(1 - p_{(j)}(c_s))}{-e^{\gamma_1} - e^{\gamma_2} \ldots - e^{\gamma_s}} \right]}{\theta} \right\}. \qquad (4.3)$$

On analysing the responses with the model described in equation (4.2), parameter estimates will be obtained. By replacing $p_{(j)}(c_s)$ in equation (4.3) with the TTL and including all the parameter estimates, the estimate of the TD associated with the TTL will be obtained. The derivation of rearrangement (4.3) is shown in Appendix 1.

4.2 Prior Information

Prior information is incorporated through the use of pseudo-data. Independent Beta distributions are placed on the conditional probabilities $\pi_{(j)1} \sim Beta(t_{(j)1}, q_{(j)1})$ such that the mean value of $\pi_{(j)1}$ is $\frac{t_{(j)1}}{t_{(j)1} + q_{(j)1}} = \frac{t_{(j)1}}{n_{(j)1}}$. The parameters $t_{(j)1}, n_{(j)1}$ are calculated based on prior opinion for $\pi_{(j)1}$, i.e. the target dose is set to correspond to TTL=20% during cycle 1, by setting $n_{(j)1} = 3$ and $\pi_{(j)1}$=0.2 then

$t_{(j)1} = \pi_{(j)1} \times n_{(j)1} = 0.6$.

Prior belief is then placed on $\pi_{(j)l}$ for cycles $l, (l = 2, \ldots, s)$. Chapter 3 suggests imposing the property of $\pi_{(j)l}$ halving for subsequent cycles, i.e. if $\pi_{(j)1}$=0.2, then $\pi_{(j)2}$=0.1, $\pi_{(j)3}$=0.05 etc.. Beta distributions are then placed on $\pi_{(j)l}$ for $l, (l = 2, \ldots, s)$

, however these are dependent on the distributions for the preceding cycles. The parameter $n_{(j)l}$ is the number of patients surviving cycle $l-1$ without a toxicity: $n_{(j)l} = n_{(j)l-1} - t_{(j)l-1}$,and therefore: $t_{(j)l} = \pi_{(j)l} \times n_{(j)l}$. The distributions are independent between dose levels however.

Pseudo-data is required for the lowest and the highest possible dose levels but also for each of the cycles of therapy. The lowest dose level is used to correspond to the TD to ensure the lowest dose is administered to the first cohort, and the highest dose corresponds to a toxic dose to ensure the procedure does not escalate too quickly. The pseudo-data needs to be conditional on that for the previous cycle dependant on prior belief as to the occurrence of toxicity with time. Table 4-1 demonstrates a generic example based 3 cycles of therapy.

Dose, $d_{(j)}$	$n_{(j)l}$	$\pi_{(j)l}$	$t_{(j)l} = n_{(j)l} \times \pi_{(j)l}$
$d_{(1)}$, cycle 1	$n_{(1)1}$	$\pi_{(1)1}$	$t_{(1)1}$
$d_{(1)}$, cycle 2	$n_{(1)2} = n_{(1)1} - t_{(1)1}$	$\pi_{(1)2}$	$t_{(1)2}$
$d_{(1)}$, cycle 3	$n_{(1)3} = n_{(1)2} - t_{(1)2}$	$\pi_{(1)3}$	$t_{(1)3}$
$d_{(k)}$, cycle 1	$n_{(k)1}$	$\pi_{(k)1}$	$t_{(k)1}$
$d_{(k)}$, cycle 2	$n_{(k)2} = n_{(k)1} - t_{(k)1}$	$\pi_{(k)2}$	$t_{(k)2}$
$d_{(k)}$, cycle 3	$n_{(k)3} = n_{(k)2} - t_{(k)2}$	$\pi_{(k)3}$	$t_{(k)3}$

Table 4- 1: Pseudo-data for $d_{(1)}$ and $d_{(k)}$, conditional on previous cycles of therapy.

The lowest dose is usually set to correspond to the TTL and the highest dose is set to correspond to a higher toxicity level. This will ensure that the lowest dose will be administered to the first cohort (when using the patient gain) and the procedure will not escalate to high doses too fast.

These prior distributions are chosen as such since the Beta distribution is easily conjugate with the Binomial likelihood obtained from the binary data and aids to the

simplicity of the procedure since more complicated approaches at eliciting prior distributions are not required.

4.3 Gain Function

The dose to be administered to the next cohort of patients depends on the gain function used in the procedure. The patient gain function would select the dose level $d_{(J)}$ that has closest $p_{(J)}(c_s)$ to the TTL and is defined below.

$$g_{i(J)} = \frac{1}{\left(TTL - p_{i(J)}(c_s)\right)^2}.$$

Where $p_{i(J)}(c_s)$ is the estimate of the probability of a DLT in the first s cycles for dose level $d_{(J)}$ after i patients.

The variance gain function would select the set of dose levels J that reduces the asymptotic variance of the estimated log transformed TD the most and is defined as in equation (4.4).

$$g_{i(J)} = \left(\frac{1}{\text{var}\left(\log\left(TD_{i,TTL}^{(+J)}\right)\right)} \right) \qquad (4.4)$$

In this setting, TD_{TTL} is the dose that is believed to correspond exactly to the TTL and $TD_{i,TTL}^{(+J)}$ is the expected estimate of the TD_{TTL} after i patients when incorporating the set of dose levels J for the next cohort of patients. The variance is calculated as follows:

$$\text{var}\left(\log\left(TD_{i,TTL}^{(+J)}\right)\right) = \nabla\left(\log\left(TD_{i,TTL}^{(+J)}\right)\right)^T I_E^{-1}(\theta, \gamma_1, \gamma_2, ..., \gamma_s)\nabla\left(\log\left(TD_{i,TTL}^{(+J)}\right)\right),$$

where I_E^{-1} is the Expected Information Matrix found from twice differentiating the log-likelihood with respect to each parameter and taking the expectation of each element

of the matrix. The parameters are replaced with the estimated values after the i^{th} observation and the set of doses that produce the smallest variance of the estimated TD_{TTL} (increases the gain, $g_{i(J)}$) are allocated to the next cohort. The set of doses J can consist of different doses if it is found that a combination of doses to administer to the next cohort reduces the variance the most.

Regardless of the gain-function, the lowest dose is always administered to the first cohort. When the patient gain is utilised, the lowest dose will be selected by the gain function since the prior is set as such so that the lowest dose corresponds to the TTL. For the variance gain, the lowest dose is administered regardless of the outcome of the gain function since it is typical to start dose-escalation procedures with the lowest dose. The observations from the first cohort are then combined with the pseudo-data and analysed to obtain posterior estimates for all parameters. The gain function is then applied again to choose further doses to administer.

4.4 Escalation Features

In order to carry out the trial efficiently, it should only be continued until a precise enough estimate of the log transformed TD_{TTL} is produced. The precision is tested by computing the ratio of the exponentiated limits of the asymptotic credible interval (CI) of the estimate of log(TD) after each new set of observations is accrued. Other methods of testing the precision could be used, such as looking at the standard deviation of the linear predictor. The asymptotic CI does however allow some additional inference as to the range of the estimates at the current time, which simpler statistics may not. In order to calculate this CI after i observations, the asymptotic variance of the estimate of $\log(TD_{TTL})$ after i observations is required, which is defined as;

$$\mathrm{var}\left(\log\left(TD_{i,TTL}\right)\right) = \nabla\left(\log\left(TD_{i,TTL}\right)\right)^{T} I_{O}^{-1}(\theta,\gamma_1,\gamma_2,...,\gamma_s)\nabla\left(\log\left(TD_{i,TTL}\right)\right).$$

Here, $I_{O}^{-1}(\theta,\gamma_1,\gamma_2,...,\gamma_s)$ is the inverted observed information matrix, found by twice differentiating the log-likelihood (with respect to each of the parameters) and taking the negative values of each derivative. $\nabla\left(\log\left(TD_{TTL}\right)\right)$ is the first derivative vector of $\log\left(TD_{TTL}\right)$ with respect to each of the parameters. For 3 cycles, the derivation of $I_{O}^{-1}(\theta,\gamma_1,\gamma_2,...,\gamma_s)$ is shown in Appendix 2. The observed information matrix for 3 cycles is displayed.

$I_{O}(\theta,\gamma_1,\gamma_2,...,\gamma_s) =$

$$\begin{pmatrix} -\sum_{j=1}^{k} R_1 & 0 & 0 & -\sum_{j=1}^{k} R_1(\log d_{(j)}) \\ 0 & -\sum_{j=1}^{k} R_2 & 0 & -\sum_{j=1}^{k} R_2(\log d_{(j)}) \\ 0 & 0 & -\sum_{j=1}^{k} R_3 & -\sum_{j=1}^{k} R_3(\log d_{(j)}) \\ -\sum_{j=1}^{k} R_1(\log d_{(j)}) & -\sum_{j=1}^{k} R_2(\log d_{(j)}) & -\sum_{j=1}^{k} R_3(\log d_{(j)}) & -\sum_{j=1}^{k} (R_1+R_2+R_3)((\log d_{(j)})^2) \end{pmatrix}.$$

Where:

$$R_1 = \frac{n_{(j)1}\pi_{(j)1}{}^2 \log(1-\pi_{(j)1}) - t_{(j)1}\left(\log(1-\pi_{(j)1})\right)^2 - t_{(j)1}\pi_{(j)1}\log(1-\pi_{(j)1})\left\{1-\log(1-\pi_{(j)1})\right\}}{\pi_{(j)1}{}^2},$$

$$R_2 = \frac{n_{(j)2}\pi_{(j)2}{}^2 \log(1-\pi_{(j)2}) - t_{(j)2}\left(\log(1-\pi_{(j)2})\right)^2 - t_{(j)2}\pi_{(j)2}\log(1-\pi_{(j)2})\left\{1-\log(1-\pi_{(j)2})\right\}}{\pi_{(j)2}{}^2},$$

$$R_3 = \frac{n_{(j)3}\pi_{(j)3}{}^2 \log(1-\pi_{(j)3}) - t_{(j)3}\left(\log(1-\pi_{(j)3})\right)^2 - t_{(j)3}\pi_{(j)3}\log(1-\pi_{(j)3})\left\{1-\log(1-\pi_{(j)3})\right\}}{\pi_{(j)3}{}^2}.$$

The determinant required to invert the matrix is;

$$\det = \sum_{j=1}^{k} R_1 \sum_{j=1}^{k} R_2 \sum_{j=1}^{j} R_3 \sum_{j=1}^{k} (R_1 + R_2 + R_3)\left((\log d_{(j)})^2\right) - \left(\sum_{j=1}^{k} R_1 \log d_{(j)}\right)^2 \sum_{j=1}^{k} R_2 \sum_{j=1}^{k} R_3$$

$$- \left(\sum_{j=1}^{k} R_2 \log d_{(j)}\right)^2 \sum_{j=1}^{k} R_1 \sum_{j=1}^{k} R_3 - \left(\sum_{j=1}^{k} R_3 \log d_{(j)}\right)^2 \sum_{j=1}^{k} R_1 \sum_{j=1}^{k} R_2.$$

The current modal estimates (as calculated by using maximum likelihood methods but incorporating prior information) for each of the parameters are then included in the expression for the asymptotic variance in equation (4.5).

$$\frac{1}{\hat{\theta}^2 \det} \times$$

$$\left\{\left(\frac{e^{\gamma_1}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)^2\left(\frac{\det + \sum R_2 \sum R_3 \left(\sum R_1 \log d_{(j)}\right)^2}{-\sum R_1}\right) - \frac{e^{\gamma_1}e^{\gamma_2}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\left(\sum R_3 \sum R_1 \log d_{(j)} \sum R_2 \log d_{(j)}\right)\right.$$

$$- \frac{e^{\gamma_1}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\left(\sum R_2 \sum R_1 \log d_{(j)} \sum R_3 \log d_{(j)}\right) + \frac{e^{\gamma_1}\log(TD_{TTL})}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\left(\sum R_2 \sum R_3 \sum R_1 \log d_{(j)}\right) -$$

$$\frac{e^{\gamma_2}e^{\gamma_1}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\left(\sum R_3 \sum R_2 \log d_{(j)} \sum R_1 \log d_{(j)}\right) + \left(\frac{e^{\gamma_2}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)^2\left(\frac{\det + \sum R_1 \sum R_3 \left(\sum R_2 \log d_{(j)}\right)^2}{-\sum R_2}\right)$$

$$- \frac{e^{\gamma_2}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\left(\sum R_1 \sum R_2 \log d_{(j)} \sum R_3 \log d_{(j)}\right) + \frac{e^{\gamma_2}\log(TD_{TTL})}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\left(\sum R_1 \sum R_3 \sum R_2 \log d_{(j)}\right) -$$

$$\frac{e^{\gamma_3}e^{\gamma_1}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\left(\sum R_2 \sum R_3 \log d_{(j)} \sum R_1 \log d_{(j)}\right) - \frac{e^{\gamma_3}e^{\gamma_2}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\left(\sum R_1 \sum R_3 \log d_{(j)} \sum R_2 \log d_{(j)}\right)$$

$$+ \left(\frac{e^{\gamma_3}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)^2\left(\frac{\det + \sum R_1 \sum R_2 \left(\sum R_3 \log d_{(j)}\right)^2}{-\sum R_3}\right) + \frac{e^{\gamma_3}\log(TD_{TTL})}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\left(\sum R_1 \sum R_2 \sum R_3 \log d_{(j)}\right) +$$

$$\frac{\log(TD_{TTL})e^{\gamma_1}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\left(\sum R_2 \sum R_3 \sum R_1 \log d_{(j)}\right) + \frac{\log(TD_{TTL})e^{\gamma_2}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\left(\sum R_1 \sum R_3 \sum R_2 \log d_{(j)}\right) +$$

$$\left.\frac{\log(TD_{TTL})e^{\gamma_3}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\left(\sum R_1 \sum R_2 \sum R_3 \log d_{(j)}\right) - \left(\log(TD_{TTL})\right)^2 \sum R_1 \sum R_2 \sum R_3\right\}. \qquad (4.5)$$

The extension of this variance to s cycles is shown in Appendix 4.

The square root of this variance is used to calculate the 95% CI for the estimate of $\log(TD_{TTL})$ using the following method:

$$\left[\log\left(TD_{TTL}\right) \pm 1.96\sqrt{\text{var}\left(\log\left(TD_{TTL}\right)\right)}\right],$$

where 1.96 is the 97.5th percentile of the standard Normal distribution.

These limits are then exponentiated to achieve a CI on the real scale which will be strictly greater than 0. When the ratio of these limits fall below a pre-defined threshold, the estimate of the TD_{TTL} is deemed accurate enough and the trial can terminate.

The difference between the asymptotic variance used for the precision criterion and for the variance gain is that for the precision criterion, only observed values from the trial so far are used within the expression for the variance. This then gives a variance value associated with what has occurred in the study so far and the current estimate of the TD_{TTL}. For the variance gain, the additional expected amount of observations for all possible sets of dose combinations are incorporated into the expression for the variance. This variance value therefore predicts which set of dose combinations will reduce the variance the most after the next cohort, so selects that combination accordingly for administration.

4.5 Proportional Odds Decision Procedure

An intermediate procedure could be considered to generalise the LRDP to include a proportional odds model which can then account for the occurrence of an event in different cycles, or indeed the occurrence of no event.

The probability of category 1 occurring, which is probability of a DLT occurring in cycle 1, is denoted as $p_{(J)1}$, the probability of category 2 (a DLT occurs in cycle 2), is $p_{(J)2}$, category 3 is $p_{(J)3}$ (a DLT occurs in cycle 3) and the probability of category 4,

which is the probability that a DLT does not occur in cycle 1, 2 or 3, is $p_{(j)4}$. $p_{(j)4}$ is directly dependent on the other categories, so can be defined as $1-p_{(j)1}-p_{(j)2}-p_{(j)3}$.

The proportional odds model uses the cumulative relationship of an event occurring in the first l cycles at dose level (j) as defined below:

$$P(T_{(j)} < c_l) = Q_{(j)l} = \frac{\exp\left(\alpha_l + \beta \log\left(d_{(j)}\right)\right)}{1+\exp\left(\alpha_l + \beta \log\left(d_{(j)}\right)\right)}.$$

Therefore, the probability of each category occurring can be defined as:

$$\begin{aligned} p_{(j)1} &= Q_{(j)1} \\ p_{(j)2} &= Q_{(j)2} - Q_{(j)1} \\ p_{(j)3} &= Q_{(j)3} - Q_{(j)2} \\ p_{(j)4} &= 1 - Q_{(j)3}. \end{aligned}$$

Furthermore, unlike the ICS model, dropout between cycles is not accounted for here. Therefore further categories need to be defined for the event that a patient drops out after cycle 1 or after cycle 2. P(Surviving cycle 1 then drop out)= $1-Q_{(j)1}$, P(Surviving cycle 2 then drop out)= $1-Q_{(j)2}$. The likelihood for this model must incorporate all eventualities and is shown below.

$$\begin{aligned} \prod_{j=1}^{k}\prod_{q=1}^{4} p_{(j)q}^{\,t_{(j)q}} &= \prod_{j=1}^{k} p_{(j)1}^{\,t_{(j)1}} p_{(j)2}^{\,t_{(j)2}} p_{(j)3}^{\,t_{(j)3}} (1-p_{(j)1}-p_{(j)2}-p_{(j)3})^{t_{(j)4}} (1-p_{(j)1})^{t_{(j)5}} (1-p_{(j)1}-p_{(j)2})^{t_{(j)6}} \\ &= \prod_{j=1}^{k} Q_{(j)1}^{\,t_{(j)1}} (Q_{(j)2}-Q_{(j)1})^{t_{(j)2}} (Q_{(j)3}-Q_{(j)3})^{t_{(j)3}} (1-Q_{(j)3})^{t_{(j)4}} (1-Q_{(j)1})^{t_{(j)5}} (1-Q_{(j)2})^{t_{(j)6}}. \end{aligned}$$

$t_{(j)4}$, $t_{(j)5}$ and $t_{(j)6}$ are the number of occurrences of patients surviving all cycles of therapy, dropping out after surviving one cycle of therapy, and dropping out after surviving two cycles of therapy respectively.

This likelihood is not of a standard form and therefore the existing procedure, which makes use of the fact that the Binomial likelihood is conjugate with a Beta prior distribution, is now no longer an easy one to use. The combination of a non-standard likelihood with any type of prior distribution would be a much more complex process and therefore not as attractive as the ICSDP that has been developed.

This approach will not be investigated further and focus will remain on using the ICS model. The ICS model naturally accounts for dropouts between cycles and produces a likelihood of a standard form which can then be combined with a conjugate prior to produce a tractable posterior distribution for the probability of a DLT.

5. The Interval-Censored Survival Decision Procedure: A Simulation Study

5.1 A Comparison of 3 Designs

The first simulation study compares the LRDP (as described in Chapter 2) for the toxicities observed in the first cycle of therapy (LRDP1), the same LRDP for the toxicities observed in the whole trial and the ICSDP (as described in Chapter 4) for the toxicities observed in every cycle of therapy for the whole trial. In this instance the "whole trial" indicates 3 cycles of therapy. As has been shown in Chapter 3, the majority of events occur in the first 3 cycles, so it can be deemed suitable to focus investigation on these cycles. The LRDP3 is then the second design used for comparison.

Each of the models used in the different procedures consist of intercepts (one intercept for the LRDPs and an intercept for every 'interval' or cycle in the ICSDP) and the coefficient of log(dose). No other covariates (such as patient characteristics e.g. gender, age, biomarkers etc.) are included in these models.

5.2 Data Generation Scenarios

5.2.1 Introduction

In order to investigate the different procedures effectively, use of the procedures under different scenarios will be investigated. When the analysis model matches the scenario

used for data generation, the results should be good compared to the true results, whereas when there is a discrepancy between analysis and generation methods, the results should be than observed when the methods match. Since the procedures to be compared consist of different analysis models, 3 different data generation methods will be adopted. First, the data will be generated by the proportional odds (PO) model, which is a generalized version of the LR model, i.e. the LR model is the PO model with 2 categories. Second, via the ICS model and third, generation from an independent model, which is the PO model but with dose as a covariate rather than log(dose). Although the assumptions of this third generation model match those of the analysis model in the LRDP, due to the different scale of dose, the models are different since the parameters estimated are different. When log(dose) is incorporated, the parameters are log odds ratios (log-OR), whereas when dose is used, the parameters are odds ratios (OR).

5.2.2 Proportional Odds Model

Simulations performed by Zhou & Whitehead [16] based on data from Ferry et. al [19], use the Logistic Regression model with the logit link function as in equation (5.1) to generate the data:

$$\log\left(\frac{p_j(c_1)}{1-p_j(c_1)}\right) = \alpha_1 + \beta \log(d_j). \qquad (5.1)$$

Here $p_j(c_1)$ is the probability of DLT for the dose d_j on the continuous scale by the end of cycle 1. The 'standard' scenario in [16] has been adopted for the generation in

which $\alpha_1 = -11.8733$ and $\beta = 1.7767$. The dose-probability of toxicity curve is shown in Figure 5-1:

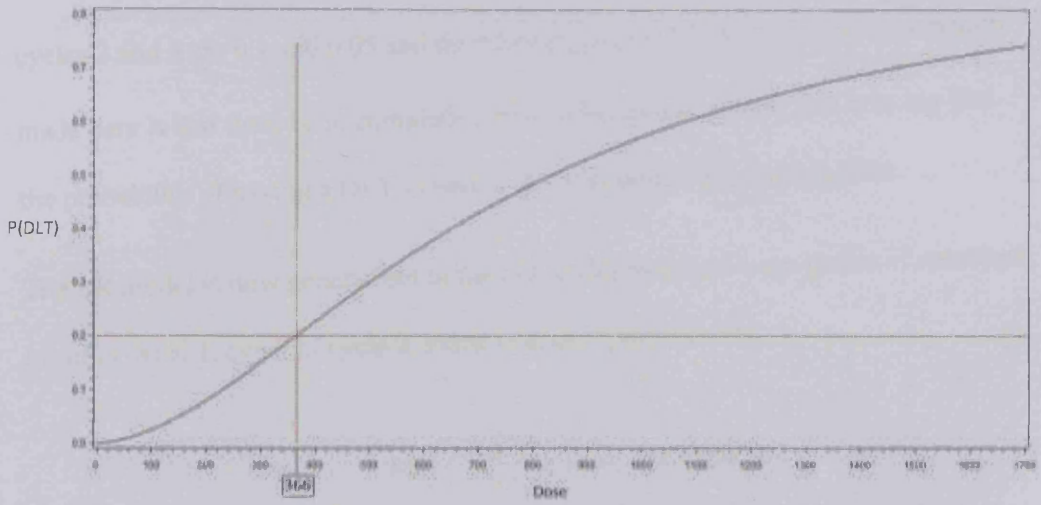


Figure 5- 1: True Dose-Response Relationship used for Simulation

This gives a TD_{20} (the target dose corresponding to a 20% chance of DLT) for the first cycle of therapy (c_1) of 366mg/m^2 as follows:

$$TD_{20} = \exp\left(\frac{\log\left(\frac{0.2}{1-0.2}\right)+11.8733}{1.7767}\right) = 366.$$

When considering multiple cycles of therapy, a higher probability of toxicity needs to be considered that incorporates the fact that the probability of toxicity will be decreasing with each cycle of therapy survived without a DLT. The overall probability of toxicity for dose d_j after 3 cycles is shown in equation (5.2):

$$p_j(c_3) = p_{j1} + p_{j2} + p_{j3} = \pi_{j1} + \pi_{j2}(1-\pi_{j1}) + \pi_{j3}(1-\pi_{j2})(1-\pi_{j1}), \qquad (5.2)$$

where p_{jl} denotes the unconditional probability of a DLT on dose d_j during cycle l and π_{jl} denotes the conditional probability of a DLT on dose d_j during cycle l. As

described in Chapter 3, at the target dose, set to be 366mg/m^2,

$\pi_{366,3} = \frac{1}{2}\pi_{366,2} = \frac{1}{4}\pi_{366,1}$. On assigning $\pi_{366,1} = 0.2$, the respective probabilities for cycles 2 and 3 are 0.1 and 0.05 and therefore $p_{366}(c_3) = 0.316$. A further assumption made here is that there is no cumulative dose effect across cycles, that is to say that the probability of having a DLT in each cycle is conditionally independent.

The LR model is now generalised to the PO model based on 4 categories of outcomes, event in cycle 1, event in cycle 2, event in cycle 3, or no event.

$$\begin{aligned} \log\left(\frac{p_j(c_1)}{1-p_j(c_1)}\right) &= \alpha_1 + \beta\log(d_j), \\ \log\left(\frac{p_j(c_2)}{1-p_j(c_2)}\right) &= \alpha_2 + \beta\log(d_j), \\ \log\left(\frac{p_j(c_3)}{1-p_j(c_3)}\right) &= \alpha_3 + \beta\log(d_j), \end{aligned} \tag{5.3}$$

where $p_j(c_l)$ is the cumulative probability of DLT up to the end of cycle l on dose d_j for $l = 1,2,3$. The final category associated with the PO model is that a DLT does not occur. The probability associated with this is simply $1 - p_j(c_3)$.

The parameters α_1 and β associated with the first equation in (5.3) are already defined. Replacing $p_j(c_3)$ in the third equation of (5.3) with 0.316, which is the TTL after 3 cycles of therapy at the target dose 366mg/m^2, and keeping the log(dose) coefficient ($\beta = 1.7767$), α_3 can be calculated.

$$TD_{31.6} = 366 = \exp\left(\frac{\log\left(\frac{0.316}{1-0.316}\right) - \alpha_3}{1.7767}\right).$$

This gives $\alpha_3 = 11.2594$. Keeping the same log(dose) coefficient is suitable since the LR model is a simplified version of the PO model (with 2 categories), therefore, extending the LR model to the PO model with more than 2 categories would require the same log-OR for the log(dose) coefficient to ensure proportional odds across categories is maintained.

The remaining intercept in equation (5.3), α_2, can be found to correspond to the probability of toxicity after 2 cycles. Given the relevant conditional probabilities, $p_{366}(c_2) = \pi_{366,1} + \pi_{366,2}(1 - \pi_{366,1}) = 0.28$. Replacing $p_j(c_2)$ with 0.28 and keeping $\beta = 1.7767$ allows the calculation of α_2 as follows:

$$TD_{28} = 366 = \exp\left(\frac{\log\left(\frac{0.28}{1-0.28}\right) - \alpha_2}{\beta}\right),$$

which gives $\alpha_2 = -11.4317$.

The values for $\alpha_1, \alpha_2, \alpha_3$ and the dose response parameter β, can be used to calculate the probability of DLT for each cycle. The overall occurrence of a DLT throughout the 3 cycles can be generated from $p_{(j)}(c_3)$ for each dose level $j = 1, \ldots, k$ through the use of a Bernoulli random variable. The possible dose levels are (60, 120, 200, 300, 420, 630, 945, 1400, 1700) as in [16, 19].

To generate the occurrence of a DLT in a specific cycle from the PO model, four categories are defined: 1 = first DLT occurs in cycle 1, 2 = first DLT occurs in cycle 2, 3 = first DLT occurs in cycle 3, 4 = no DLT occurs during the first three cycles.

The outcome category for a subject on dose $d_{(j)}$ is generated from a Uniform[0,1] distribution in the proportions $p_{(j)1}, p_{(j)2}, p_{(j)3}, 1-p_{(j)3}$ for categories 1–4 respectively.

For data generated under the PO model, the LRDPs should perform better than the ICSDP. To show the discrepancy between the data generation and the analysis models for the ICSDP, the relationship between the probability of toxicity and the conditional probabilities of toxicity for each cycle can be considered. The relationship between p_{jl}, the probability of toxicity in cycle l, and π_{jl}, the conditional probability of toxicity in cycle l given that there is no toxicity up to and including cycle $l-1$ is as follows:

$$\pi_{j1} = p_{j1},\ \pi_{j2} = \frac{p_{j2}}{1-p_j(c_1)},\ \pi_{j3} = \frac{p_{j3}}{1-p_j(c_2)},\ \ldots,\ \pi_{js} = \frac{p_{js}}{1-p_j(c_{s-1})}. \qquad (5.4)$$

The PO model with parameters $\alpha_1 = -11.8733, \alpha_2 = -11.4317, \alpha_3 = -11.2594$, $\beta = 1.7767$, the $p_j(c_3)$ can be found for some pre-specified dose levels $d_{(j)}$ for $j = 1, 60, 300, 1500$. These dose levels are chosen since the latter 3 are scaled equally. By doing this one would hope that any difference between dose levels would be consistent between dose 60 and 300, and 300 and 1500. Any deviation from this may indicate the model performs even worse when the dose increases. This implies that comparison of 300 to 60, and 1500 to 300 is equivalent. A dose of 1 will allow estimation of the intercept parameters by eliminating the log(dose) coefficient. The unconditional cycle probabilities $p_{(j)1}$, $p_{(j)2}$, $p_{(j)3}$ are obtained by calculating $p_{(j)}(c_1)$, $p_{(j)}(c_2)$ and $p_{(j)}(c_3)$ so that $p_{(j)1} = p_{(j)}(c_1)$, $p_{(j)2} = p_{(j)}(c_2) - p_{(j)}(c_1)$ and $p_{(j)3} = p_{(j)}(c_3) - p_{(j)}(c_2)$. The conditional probabilities $\pi_{(j)l}$ can then be found from equation (5.4). The intercept terms for each cycle $(\gamma_1, \gamma_2, \gamma_3)$ can be from obtained

from π_{1l} by using the complementary log-log link function

$\left(\log\left(-\log\left(1-\pi_{jl}\right)\right)=\gamma_l+\theta\log\left(d_j\right)\right)$ where $d_j=1$. The log hazard ratio (θ) is then calculated for each cycle and dose. The results from doing this are summarised in Table 5-1.

PO model: $\alpha_1 = -11.8733, \alpha_2 = -11.4317, \alpha_3 = -11.2594, \beta = 1.7767$			
d=60	d=300	d=1500	d=1
$p_{60}(c_1) = 0.0100$	$p_{300}(c_1) = 0.1494$	$p_{1500}(c_1) = 0.7450$	$p_1(c_1) = 0.000007$
$p_{60}(c_2) = 0.0154$	$p_{300}(c_2) = 0.2145$	$p_{1500}(c_2) = 0.8176$	$p_1(c_3) = 0.000011$
$p_{60}(c_3) = 0.0183$	$p_{300}(c_3) = 0.2450$	$p_{1500}(c_3) = 0.8499$	$p_1(c_3) = 0.000013$
$p_{(j)1} = p_{(j)}(c_1), p_{(j)2} = p_{(j)}(c_2) - p_{(j)}(c_1), p_{(j)3} = p_{(j)}(c_3) - p_{(j)}(c_2)$			
$p_{60,1} = 0.0100$	$p_{300,1} = 0.1494$	$p_{1500,1} = 0.7540$	$p_{1,1} = 0.000007$
$p_{60,2} = 0.0054$	$p_{300,2} = 0.0651$	$p_{1500,2} = 0.0726$	$p_{1,2} = 0.000004$
$p_{60,3} = 0.0029$	$p_{300,2} = 0.0305$	$p_{1500,3} = 0.0233$	$p_{1,3} = 0.000002$
$\pi_{j1} = p_{j1}, \pi_{j2} = \frac{p_{j2}}{1-p_{j1}}, \pi_{(j)3} = \frac{p_{(j)3}}{1-p_{(j)1}-p_{(j)2}}$			
$\pi_{60,1} = 0.0100$	$\pi_{300,1} = 0.1494$	$\pi_{1500,1} = 0.7540$	$\pi_{1,1} = 0.000007$
$\pi_{60,2} = 0.0055$	$\pi_{300,2} = 0.0765$	$\pi_{1500,2} = 0.2951$	$\pi_{1,2} = 0.000004$
$\pi_{60,3} = 0.0029$	$\pi_{300,3} = 0.0384$	$\pi_{1500,3} = 0.1344$	$\pi_{1,3} = 0.000002$
ICS model: From $\pi_{l,1}$: $\gamma_1 = -11.8696, \gamma_2 = -12.4292, \gamma_3 = -13.1224$			
Solve for θ from: $\log(-\log(1-\pi_{(j),l})) = \gamma_l + \theta \log d_{(j)}$			
l=1, θ = 1.7755	l=1, θ = 1.7617	l=1, θ = 1.6693	
l=2, θ = 1.7656	l=2, θ = 1.7354	l=2, θ = 1.5559	
l=3, θ = 1.7783	l=3, θ = 1.7082	l=3, θ = 1.5297	

Table 5- 1: Checking the simulation method does not match the analysis method.

It can be seen from Table 5-1 that θ, the log(HR), is not constant across cycles or doses.

5.2.3 Interval-Censored

In order to simulate according to an ICS model, the data must be progressively simulated. So toxicities in the first cycle are generated according to a Bernoulli model with probability of toxicity $\pi_{(j)1} = p_{(j)1}$. Of the remaining patients who do not observe a toxicity in cycle 1, the next cycle's toxicities are simulated with probability of

toxicity, $\pi_{(j)2} = \dfrac{p_{(j)2}}{1-p_{(j)1}}$, and the same for the final cycle. The remaining patients who have not experienced a toxic event at dose level $d_{(j)}$, have toxicities generated for cycle 3 with probability of toxicity, $\pi_{(j)3} = \dfrac{p_{(j)3}}{1-p_{(j)1}-p_{(j)2}}$.

Values for $\pi_{(j)l}$ are now found from the model based on the complementary log-log link function and are of the form:

$$\pi_{j1} = \exp\left(-\exp\left(\gamma_1 + \theta\log\left(d_j\right)\right)\right),\ \pi_{j2} = \exp\left(-\exp\left(\gamma_2 + \theta\log\left(d_j\right)\right)\right)$$
$$\pi_{j3} = \exp\left(-\exp\left(\gamma_3 + \theta\log\left(d_j\right)\right)\right).$$

In order to find appropriate parameter values, the previous scenario adopted from Zhou, Whitehead [16] allocates the doses 366 and 799 to a 20% and 50% chance of toxicity in cycle 1. Assuming the probability of toxicity halves at the target dose (366mg/m^2) such that $\pi_{366,3} = \frac{1}{2}\pi_{366,2} = \frac{1}{4}\pi_{366,1}$, 4 equations can be set up to provide the parameter values. These are:

$$\begin{aligned}
\log\left(-\log\left(1-\pi_{366,1}\right)\right) &= \log\left(-\log\left(1-0.2\right)\right) = \gamma_1 + \theta\log\left(366\right),\\
\log\left(-\log\left(1-\pi_{366,2}\right)\right) &= \log\left(-\log\left(1-0.1\right)\right) = \gamma_2 + \theta\log\left(366\right),\\
\log\left(-\log\left(1-\pi_{366,3}\right)\right) &= \log\left(-\log\left(1-0.05\right)\right) = \gamma_3 + \theta\log\left(366\right),\\
\log\left(-\log\left(1-\pi_{799,1}\right)\right) &= \log\left(-\log\left(1-0.5\right)\right) = \gamma_1 + \theta\log\left(799\right).
\end{aligned} \tag{5.5}$$

The values obtained from solving these equations are $\gamma_1 = -10.0694$, $\gamma_2 = -10.8198$, $\gamma_3 = -11.5396$ and $\theta = 1.4518$.

It is again important to show that the data simulated from this model does not have a constant log odds ratio (log OR) across doses (as it would for the logistic regression model) so that one can see how the logistic regression model works when the model

assumption is incorrect. The approach utilised in section 5.2.2 is applied in reverse here. First the $\pi_{(j)l}$ are calculated for the dose levels (1,60,300,1500). The unconditional probabilities $p_{(j)l}$ are then computed from rearranging equation (5.4). Finally the $p_{(j)}(c_l)$ are calculated as in equation (5.2). The intercept terms for α_1 and α_3 are calculated from the logit link function $\log\left(\frac{p_{(j)}(c_l)}{1-p_{(j)}(c_l)}\right) = \alpha_l + \beta\log(d_{(j)})$. The intercept term α_2 is not of as much interest since the LRDP1 and LRDP3 analyse results after 1 and 3 cycles, so the log OR (β) will only be estimated with α_1 and α_3. The results are summarised below.

ICS model: $\gamma_1 = -10.0694, \gamma_2 = -10.8198, \gamma_3 = -11.5396, \theta = 1.4518$			
d=60	d=300	d=1500	d=1
$\pi_{60,1} = 0.0160$	$\pi_{300,1} = 0.1540$	$\pi_{1500,1} = 0.8227$	$\pi_{1,1} = 0.000042$
$\pi_{60,2} = 0.0076$	$\pi_{300,2} = 0.0759$	$\pi_{1500,2} = 0.5581$	$\pi_{1,2} = 0.000020$
$\pi_{60,3} = 0.0037$	$\pi_{300,3} = 0.0377$	$\pi_{1500,3} = 0.3281$	$\pi_{1,3} = 0.000009$
$\pi_{j1} = p_{j1},\ \pi_{j2} = \frac{p_{j2}}{1-p_{j1}},\ \pi_{j3} = \frac{p_{j3}}{1-p_{j2}-p_{j1}}$			
$p_{60,1} = 0.0160$	$p_{300,1} = 0.1540$	$p_{1500,1} = 0.8227$	$p_{1,1} = 0.000042$
$p_{60,2} = 0.0075$	$p_{300,2} = 0.0897$	$p_{1500,2} = 0.0990$	$p_{1,2} = 0.000020$
$p_{60,3} = 0.0037$	$p_{300,2} = 0.0482$	$p_{1500,3} = 0.0257$	$p_{1,3} = 0.000009$
$p_{(j)}(c_1) = p_{(j)1},\ p_{(j)}(c_2) = p_{(j)2} + p_{(j)1},\ p_{(j)}(c_3) = p_{(j)3} + p_{(j)2} + p_{(j)1}$			
$p_{60}(c_1) = 0.0160$	$p_{300}(c_1) = 0.1540$	$p_{1500}(c_1) = 0.8227$	$p_1(c_1) = 0.000042$
$p_{60}(c_1) = 0.0235$	$p_{300}(c_2) = 0.2437$	$p_{1500}(c_2) = 0.9217$	$p_1(c_2) = 0.000062$
$p_{60}(c_3) = 0.0271$	$p_{300}(c_3) = 0.2919$	$p_{1500}(c_3) = 0.9474$	$p_1(c_3) = 0.000071$

Table 5- 2: Checking the simulation method does not match the analysis method.

PO model: From $p_1(c_1), p_1(c_3)$: $\alpha_1 = -10.0778, \alpha_3 = -9.5528$			
Solve for β from: $\log\left(p_{(j)}(c_l)/1-p_{(j)}(c_l)\right) = \alpha_l + \beta \log d_{(j)}$			
$l = 1, \beta = 1.4554$ $l = 3, \beta = 1.4586$	$l = 1, \beta = 1.4682$ $l = 3, \beta = 1.5195$	$l = 1, \beta = 1.5878$ $l = 3, \beta = 1.7015$	

Table 5-2 cont.: Checking the simulation method does not match the analysis method.

As can be seen in Table 5-2, the log OR is not constant across doses or cycles.

5.2.4 Proportional Odds with dose as covariate

A further model used to check for robustness is the proportional odds model again, but this time using dose as a covariate rather than log-dose.

The generation of events according to the proportional odds model with dose as a covariate rather than log-dose is similar to section 5.2.2. The doses 366 and 799mg/m^2 are set to correspond to probabilities of toxicity 0.2 and 0.5 respectively for the first cycle and the conditional probabilities halve at the dose 366 for successive cycles. Therefore the following equations allow solving for the intercept, ε_1, and the dose response parameter ϕ:

$$\log\left(\frac{p_{366}(c_1)}{1-p_{366}(c_1)}\right) = \log\left(\frac{0.2}{1-0.2}\right) = \varepsilon_1 + 366\phi,$$
$$\log\left(\frac{p_{799}(c_1)}{1-p_{799}(c_1)}\right) = \log\left(\frac{0.5}{1-0.5}\right) = \varepsilon_1 + 799\phi.$$

These can be solved to give $\varepsilon_1 = -2.5575$ and $\phi = 0.0032$. The intercepts for the cumulative probability of toxicity up to cycle 2 and 3 are also found from including the value of ϕ in the following equations:

$$\log\left(\frac{p_{366}(c_2)}{1-p_{366}(c_2)}\right)=\log\left(\frac{0.28}{1-0.28}\right)=\varepsilon_2+366\times(0.0032),$$

$$\log\left(\frac{p_{366}(c_3)}{1-p_{366}(c_3)}\right)=\log\left(\frac{0.316}{1-0.316}\right)=\varepsilon_3+366\times(0.0032).$$

This gives $\varepsilon_2=-2.1157$ and $\varepsilon_3=-1.9434$.

To check that the interpretation of the parameter values is not consistent with either of the analysis methods, the $p_{(j)}(c_l)$ for the doses (1,60,300,1500) are calculated from the values for $\varepsilon_1,\varepsilon_2,\varepsilon_3,\phi$. First these can be used to calculate the intercept terms α_1,α_3 associated with the PO model with log(dose) as a covariate, and then the log OR β can be calculated associated with cycles 1 and 3, and across dose levels. The unconditional cycle probabilities $p_{(j)l}$ are also calculated as in equation (5.2) which are then used to calculate the conditional probabilities $\pi_{(j)l}$ as in equation (5.4). These are then used to calculate the intercept terms $\gamma_1,\gamma_2,\gamma_3$ for the ICS model which are then used to calculate θ for each cycle and dose level.

Table 5-3 shows how the estimates change for the different models.

PO with dose: $\varepsilon_1=-2.5575, \varepsilon_2=-2.1157, \varepsilon_3=-1.9434, \phi=0.0032$			
d=60	d=300	d=1500	d=1
$p_{60}(c_1)=0.0859$ $p_{60}(c_2)=0.1274$ $p_{60}(c_3)=0.1479$	$p_{300}(c_1)=0.1684$ $p_{300}(c_2)=0.2394$ $p_{300}(c_3)=0.2722$	$p_{1500}(c_1)=0.9041$ $p_{1500}(c_2)=0.9361$ $p_{1500}(c_3)=0.9457$	$p_1(c_1)=0.0722$ $p_1(c_2)=0.1079$ $p_1(c_3)=0.1256$
PO with log(dose): From $p_1(c_1)$, $p_2(c_2), p_1(c_3)$: $\alpha_1=-2.5534, \alpha_2=-2.1124, \alpha_3=-1.9404$			
Solve for β from: $\log\left(p_{(j)}(c_l)\Big/1-p_{(j)}(c_l)\right)=\alpha_l+\beta\log d_{(j)}$			

Table 5- 3: Checking the simulation method does not match the analysis methods.

$l = 1, \beta = 0.0461$ $l = 3, \beta = 0.0462$	$l = 1, \beta = 0.1677$ $l = 3\beta = 0.1678$	$l = 1, \beta = 0.6560$ $l = 3, \beta = 0.6560$	
$p_{(j)1} = p_{(j)}(c_1), p_{(j)2} = p_{(j)}(c_2) - p_{(j)}(c_1), p_{(j)3} = p_{(j)}(c_3) - p_{(j)}(c_2)$			
$p_{60,1} = 0.0859$ $p_{60,2} = 0.0415$ $p_{60,3} = 0.0205$	$p_{300,1} = 0.1684$ $p_{300,2} = 0.0710$ $p_{300,3} = 0.0328$	$p_{1500,1} = 0.9041$ $p_{1500,2} = 0.0320$ $p_{1500,3} = 0.0096$	$p_{1,1} = 0.0722$ $p_{1,2} = 0.0357$ $p_{1,3} = 0.0177$
$\pi_{j1} = p_{j1}, \pi_{j2} = \frac{p_{j2}}{1-p_{j1}}, \pi_{(j)3} = \frac{p_{(j)3}}{1-p_{(j)1}-p_{(j)2}}$			
$\pi_{60,1} = 0.0859$ $\pi_{60,2} = 0.0454$ $\pi_{60,3} = 0.0234$	$\pi_{300,1} = 0.1684$ $\pi_{300,2} = 0.0854$ $\pi_{300,3} = 0.0431$	$\pi_{1500,1} = 0.9041$ $\pi_{1500,2} = 0.3337$ $\pi_{1500,3} = 0.1502$	$\pi_{1,1} = 0.0722$ $\pi_{1,2} = 0.0385$ $\pi_{1,3} = 0.0198$
Interval-Censored: From $\pi_{1,l}$: $\gamma_1 = -2.5911, \gamma_2 = -3.2375, \gamma_3 = -3.9121$			
Solve for θ from: $\log(-\log(1-\pi_{(j)l})) = \gamma_l + \theta \log d_{(j)}$			
$\theta = 0.0442$ $\theta = 0.0411$ $\theta = 0.0413$	$\theta = 0.1579$ $\theta = 0.1440$ $\theta = 0.1677$	$\theta = 0.4708$ $\theta = 0.3194$ $\theta = 0.2867$	

Table 5-3 cont.: Checking the simulation method does not match the analysis methods.

The values for β are generally the same within dose levels across cycles, but are different across dose levels. The values for θ are different across dose and cycles, with the difference across cycles increasing with dose.

5.2.5 Differences

Figures 5-2 to 5-6 show graphically the differences between data generated by one model and analysed by another. Each figure shows the dose-response curve for 1000 generated trials of 60 patients for each of the dose levels (60, 120, 200, 300, 420, 630, 945, 1400 and 1700) over 3 cycles. All of the generated data is then analysed by each possible models to estimate P(DLT) for doses on a continuous scale.

The dashed line represents the proportional odds model, either with 1 cycle or 3 cycles (defined on the graph). The solid line depicts the Interval-Censored Survival model and the dotted line displays the proportional odds model with dose as a covariate.

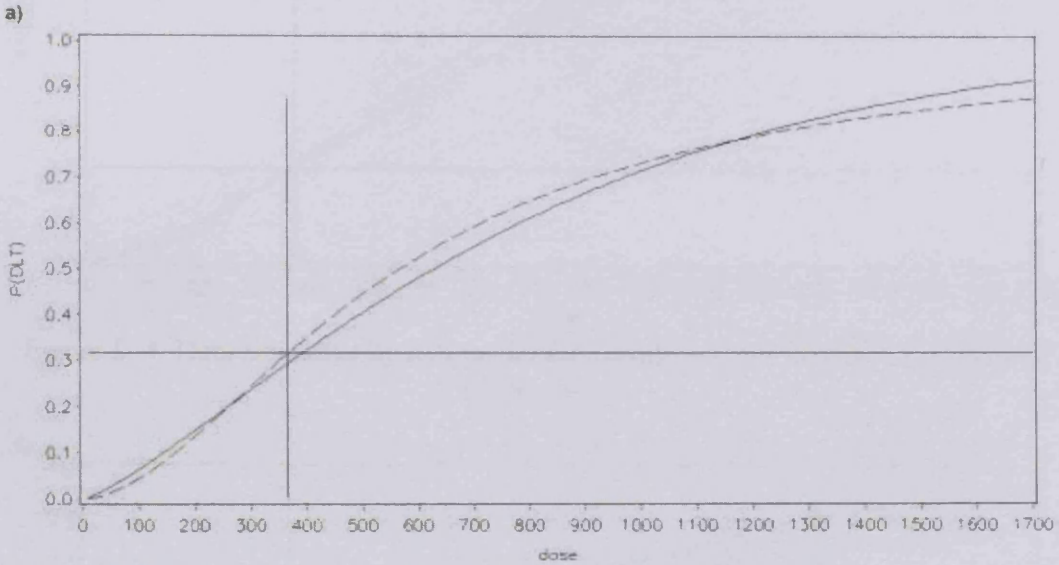


Figure 5- 2: Data simulated by PO model with log dose for 3 cycles – – – , analysed by ICS model for 3 cycles ——.

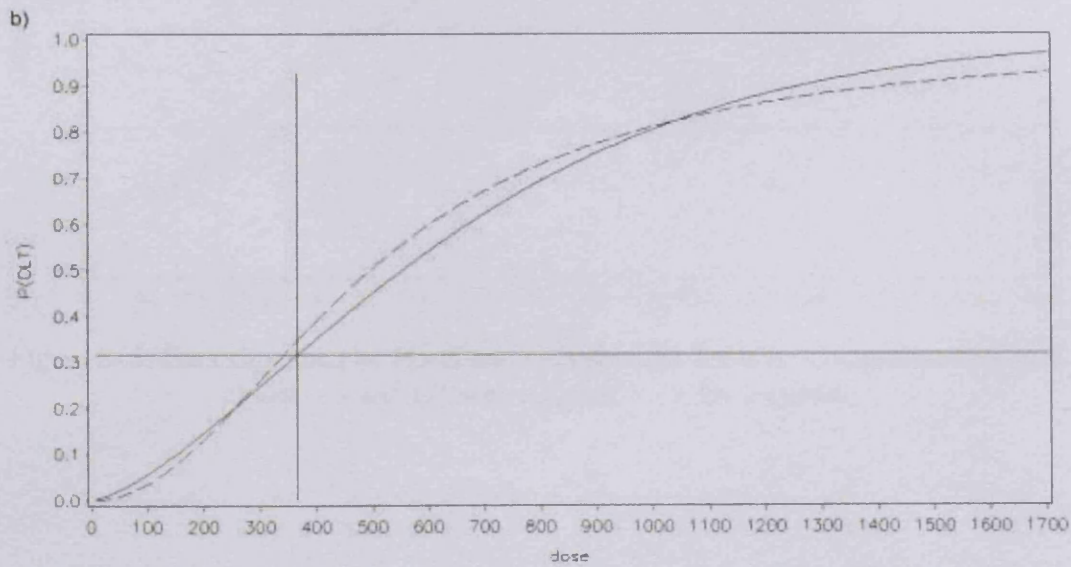


Figure 5- 3: Data simulated by ICS model for 3 cycles ——, analysed by LR model with log dose for 3 cycles – – –.

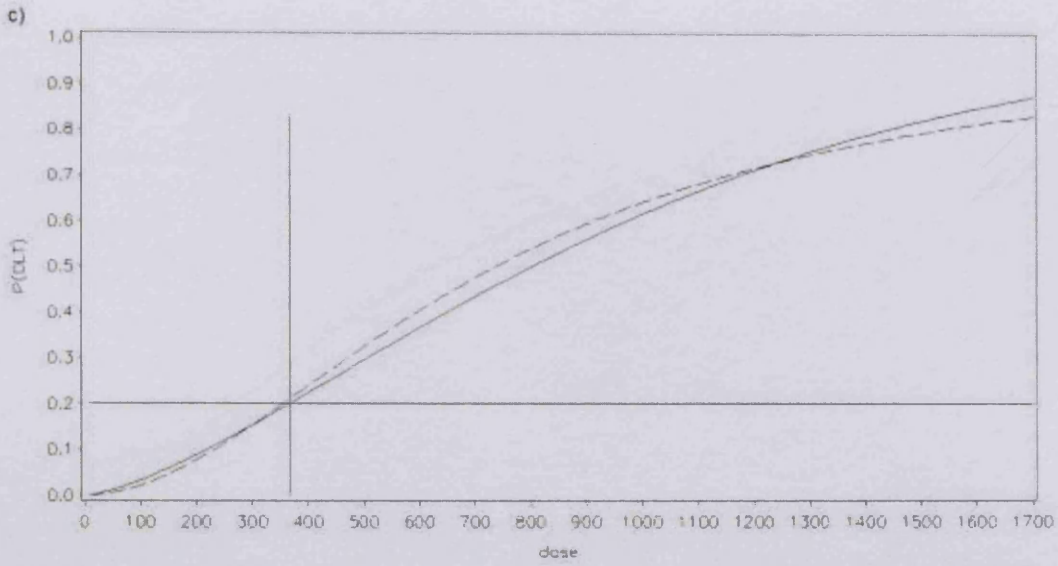


Figure 5- 4: Data simulated by ICS model for 1 cycle —— and analysed by LR for 1 cycle – – –.

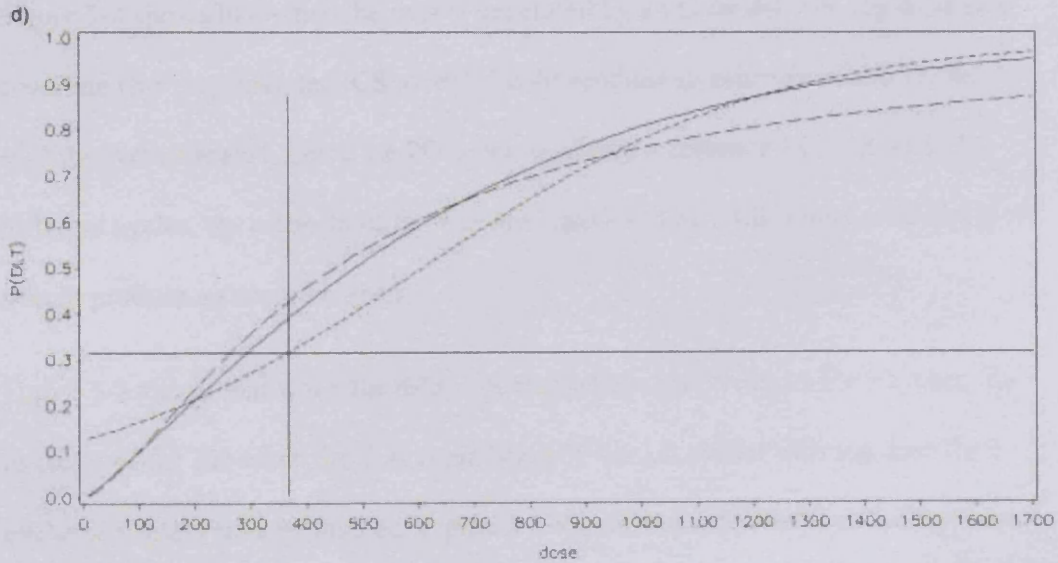


Figure 5- 5: Data simulated by PO model with dose for 3 cycles - - -, analysed by ICS model —— and LR with log dose – – – for 3 cycles.

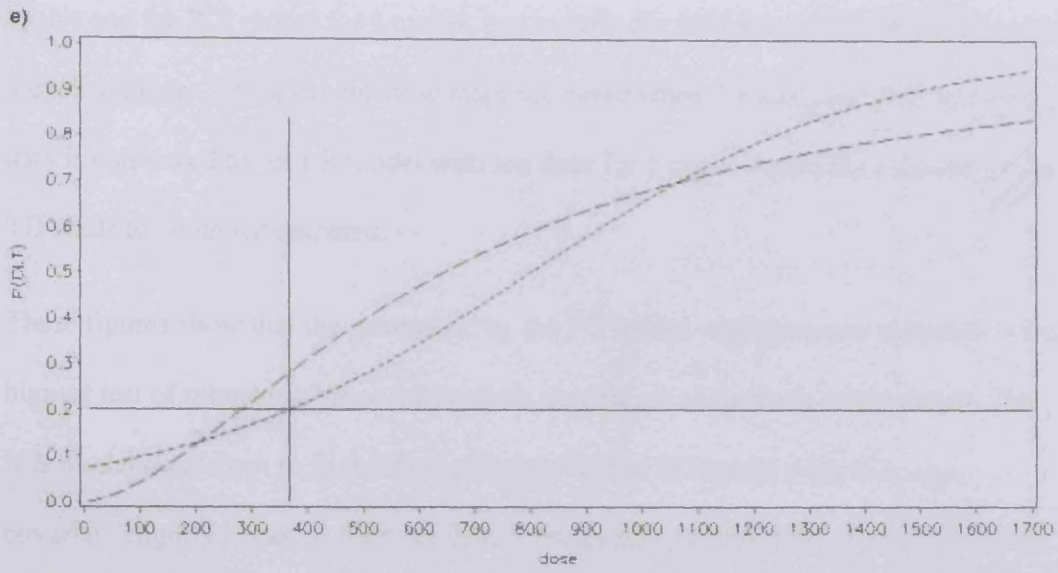


Figure 5- 6: Data simulated by PO model with dose - - - for 1 cycle, analysed by LR for 1 cycle – – –.

Figure 5-2 shows that when the data is generated by a PO model with log dose as a covariate (for 3 cycles), the ICS model should produce an estimate of the TD that is slightly overestimated. Since the PO model assumes a common log odds ratio for different cycles, the estimate of the TD when analysed by a LR model with one cycle should produce an accurate result.

Figure 5-3 shows that when the data is generated by an ICS model for 3 cycles, the estimate of the TD when the data is analysed by the LR model with log dose for 3 cycles is slightly underestimated. Figure 5-4 shows that when the first cycle (when generated by an ICS model) is analysed by the LR model with log dose for 1 cycle, the estimates are very accurate. This is because the ICS model generates sequentially with a probability of DLT being 20% for the TD in cycle 1. This is the same probability under investigation for the true LR model with 1 cycle.

Figure 5-5 shows that when the generation method is a PO model with dose as a covariate for 3 cycles, the analysis methods are both incorrect and the estimates of the TD are severely underestimated when analysed by the LR model with log dose for 3

cycles and the ICS model for 3 cycles, particularly for the LR model with log dose for 3 cycles. Figure 5-6 shows the dose response curve when the analysis of the generated data is conducted by an LR model with log dose for 1 cycle. Again the estimate of the TD tends to be underestimated.

These figures show that the generation by the PO model with dose as a covariate is the biggest test of robustness since the analysis models are so severely mismatched. The ICS model does seem to fit the data generated by the PO model with dose as a covariate slightly better, at least the dose corresponding to the TTL of 31.6% is closer to 366mg/m^2 when analysed by the ICS model rather than the PO model with log(dose). This is confirmed by Table 5-3, where the value for θ, the log-HR as calculated for the ICS model, is closer to the true value of the log-OR $(\phi = 0.0032)$ at all dose levels and cycles than the log-OR is as calculated for the PO model. This suggests there is less change when the parameter inference changes from log-OR to log-HR, than vice versa. While another model could be investigated to use for the data generation that is equally biased for analyses by both the ICS model and PO model, the data generation scenarios chosen cover a variety of biases which favour different analysis models in different circumstances. Therefore acknowledging the expected differences within the results from the data generation scenarios adopted should give a wide enough overview of how the procedures adopting different analysis models fare under model misspecification.

5.3 Pseudo-data Prior Information

Prior information is created for the LRDP1 based on the pseudo data used in Zhou, Whitehead [16] as shown in Table 5-4.

Procedure, TTL	Dose $d_{(j)}$	$p_{(j)1}{}^{0}$	$n_{(j)}{}^{0}$	$t_{(j)}{}^{0}=n_{(j)}{}^{0}p_{(j)}{}^{0}$
LRDP1 ***TTL=0.2***	d_1	0.2	3	0.6
	d_k	0.5	3	1.5

Table 5- 4: Pseudo-data for the LRDP1

The first dose level ($d_{(1)}$, 60mg/m^2) is given $t_{(1)}{}^{0}$ toxicities in $n_{(1)}{}^{0}$ patients where $t_{(1)}{}^{0}/n_{(1)}{}^{0}$ is the required probability of toxicity under investigation. In [16], $t_{(1)}{}^{0}=0.6$ and $n_{(1)}{}^{0}=3$ give rise to a probability of DLT=0.2. Further pseudo data lets the highest dose level ($d_{(k)}$, 1700mg/m^2), correspond to a TD50 so $t_{(k)}{}^{0}=1.5$ and $n_{(k)}{}^{0}=3$. Analysing this cautious prior and using the patient gain causes the lowest dose level to be assigned to the first cohort (which is usually required in Phase I Trials). This simulation study investigates a TD20 for the first cycle also so the same pseudo-data in Table 5-4 is used for the LRDP1. This pseudo data corresponds to independent Beta priors being put on the probability of DLT for dose levels $d_{(j)}, j=1,...,k$, as described in Chapter 4.2. Using this prior allows the data (which creates a Binomial likelihood) to be combined with the data easily since the prior is conjugate.

Similar pseudo data is created for the LRDP3. This simulation study investigates a TD31.6 after 3 cycles, so in this case $\frac{t_{(1)}{}^{0}}{n_{(1)}{}^{0}}=0.316$. This data is presented in Table 5-5.

Procedure, TTL	Dose $d_{(j)}$	$p^{0}_{(J)}(c_3)$	$n^{0}_{(J)}$	$t^{0}_{(J)}=p^{0}_{(J)}(c_3)n^{0}_{(J)}$
LRDP3 ***TTL=0.316***	d_1	0.316	3	0.948
	d_k	0.649	3	1.947

Table 5- 5: Pseudo-data for the LRDP3.

For the lowest dose (60mg/m^2), $t_{(1)}{}^0 = 0.948$ while $n_{(1)}{}^0$ remains at 3. For the highest dose, data from the first cycle should cause a 50% probability of toxicity. The TD50 in cycle 1 is the dose 799. Given the parameter values found for the PO model after 3 cycles ($\alpha_3 = -11.2594, \beta = 1.7767$), the dose 799 produces a probability of toxicity of 0.649. This overall probability is then assigned to the highest dose ($d_{(k)}$) of 1700mg/m^2. $n_{(k)}{}^0$ is again set to 3 and $t_{(k)}{}^0$ is found to be 1.947.

Pseudo data for the ICSDP is created in the same way but sequentially. The probability of toxicity for the first cycle for the lowest and highest doses ($j = 1, k = 60$, 1700mg/m^2) is set to be the same as for the logistic regression with one cycle corresponding to doses 366 and 799mg/m^2. For cycle 1, $n_{(1)}{}^0 = 3$, and $t_{(1)}{}^0$ is the same as for the logistic regression model for the first cycle ($t_{(1)1}{}^0 = 0.6, t_{(1)k}{}^0 = 1.5$).

According to the ICS model, $\pi_{366,3} = \frac{1}{2}\pi_{366,2} = \frac{1}{4}\pi_{366,1}$ where $\pi_{366,1} = p_{366,1} = 0.2$. These then correspond to $\pi_{(1)1}{}^0, \pi_{(1)2}{}^0$ and $\pi_{(1)3}{}^0$. Furthermore, $\pi_{799,1} = p_{799,1} = 0.5$ which then corresponds to $\pi_{(k)1}{}^0$.

$\pi_{(k)2}{}^0, \pi_{(k)3}{}^0$ can be found from the parameters obtained from the ICS model by solving the equations (5.5) with the lowest and highest dose levels included (60, 1700mg/m^2) rather than the true TD_{20}, TD_{50} (366,799mg/m^2). This produces the following parameter values $\gamma_1{}^0 = -2.8877$, $\gamma_2{}^0 = -3.6381, \gamma_3{}^0 = -4.3579$ and $\theta^0 = 0.3389$.

Based on these $\pi_{(j)l}$ the pseudo-data in Table 5-6 is developed.

Procedure, TTL	Dose $d_{(j)}$	$\pi^0_{(j)l}$	$n^0_{(j)l}$	$t^0_{(j)l} = \pi^0_{(j)l} n^0_{(j)l}$
ICSDP *TTL=0.316*	$d_{(1)}$, cycle 1	0.2	3	0.6
	$d_{(1)}$, cycle 2	0.1	2.4	0.24
	$d_{(1)}$, cycle 3	0.05	2.16	0.108
	$d_{(k)}$, cycle 1	0.5	3	1.5
	$d_{(k)}$, cycle 2	0.2791	1.5	0.41865
	$d_{(k)}$, cycle 3	0.1473	1.08135	0.1593

Table 5- 6: Pseudo-data for the ICSDP.

The number of patients in each cycle is now dependent on how many DLTs have occurred in the previous cycles, since the probabilities are conditional. So $n_{(j)l} = n_{(j),l-1} - t_{(j),l-1}$, where $n_{(j)l}$ and $t_{(j)l}$ are the number of patients and the number of toxicities on dose level $d_{(j)}$ during cycle l. So the pseudo-data for $\pi_{(j)1}$ correspond to independent Beta distributions, but the distributions $\pi_{(j)2}, \pi_{(j)3}$ are independent Beta distributions across doses, but dependent on previous cycles. These values once again result in the escalation choosing the lowest available dose (60) for the first dose to be administered when using the patient gain function. While in the ICSDP the actual total number of observations is greater than for the LRDPs, given the increased observation period for the ICSDP the increased amount of prior information is expected.

5.4 Escalation Procedure

For each of the 3 procedures, the dose escalation procedure begins by analysing the pseudo data (PROC GENMOD in SAS with the 'logit' link function for the 2 LRDPs and the 'cloglog' link function for the ICSDP). This produces parameter estimates which are used within the respective link functions (equations (2.2) and (4.2)) to produce an estimate for the target dose (TD) as in equation (2.3) or (4.3) for the LRDPs or ICSDP respectively, and $p_{(j)}(c_3)$ for each dose level $d_{(j)}$ as in equation (5.2). The dose level $d_{(j)}$ with $p_{(j)}(c_3)$ closest to the required probability of toxicity

(TTL) is administered, which is equal to the lowest available dose (60mg/m^2). This is the patient gain as defined in Chapter 2, which will be used for this simulation study.

$$g_{i(j)} = \left(\frac{1}{TTL - p_{i(j)}(c_3)} \right)^2 .$$

The dose that gives the highest gain is selected and administered to the next cohort of patients. The presence or absence of DLTs for each new patient is then observed and appended to the pseudo-data and analysed again. The model parameters are re-estimated from this analysis and used to recalculate $p_{(j)}(c_3)$ for each dose level $d_{(j)}$. In the perfect case, a new cohort will begin their first cycle of therapy at the same time the preceding cohort begins the second cycle of therapy, and the cohort prior to that; their third cycle of therapy. This is the method adopted for the simulation procedure here for simplicity, whilst in practice, it may be reasonable to assume that there is some difference between actual dates of therapy. This method is also adopted for all future simulations.

The procedure continues until one of three criteria is met. Either the safety rule is violated which is when the probability of toxicity associated with the chosen dose exceeds some threshold. For the TD20 this threshold is 0.3. The dose that corresponds to this probability of toxicity is found to be 495.659mg/m^2 from the 'standard' scenario in [16]. Using this dose to calculate the probability of toxicity after 3 cycles (with $\alpha_3 = 11.2594$ and $\beta = 1.7767$) shows that the unsafe probability of toxicity is 0.4419. Therefore the safety rule for the procedures that are looking for the TD31.6 (LRDP3 and ICSDP) will stop the trial if the dose to be administered has a probability of toxicity greater than 0.44. The other criteria are if the precision stopping rule is met

or the trial has recruited 20 cohorts of 3 patients (20 cohorts is decided to be the longest a trial should be).

The precision stopping rule is associated with the ratio of the limits of the credible interval for the TD. Once this ratio is below a certain threshold (in this case 4), the estimate of the TD is deemed accurate enough for the escalation to stop. The credible intervals are calculated after every new cohort of information is recorded. For the LRDP1 this is after every cohort has completed their first cycle of therapy. For the LRDP3 this is after a cohort has completed 3 cycles of therapy and for the ICSDP this is after every cycle for every cohort of patients. The credible intervals for the log TD are calculated by finding the asymptotic variance of the data so far as defined in equation (4.5) in Chapter 4.4. The 2.5^{th} and 97.5^{th} percentile points of the Normal distribution with the mean being the current estimate of the $\log(TD)$ and the standard deviation being the square root of the asymptotic variance of $\log(TD)$ then form the credible interval limits. This is all done on the log scale to ensure that the final (exponentiated) estimates are all positive. Once the upper and lower limits of the credible interval are found they are exponentiated and the ratio of the two is taken. This ratio (R) is the value that is used for the stopping rule, if R is below a certain threshold (R=4) the escalation is deemed accurate enough to stop. 4 was selected after some investigation as it provided an estimate of the TD for the LRDP1 within a clinically relevant range ($366 \pm (0.3 * 366)$) on average over 80% of the time and produced these results on average within 17 cohorts, which is less than the maximum of 20 cohorts.

5.5 Results

Once the trial has completed, the estimate of the *TD* is recorded along with the length of the trial, defined as the cohort number of the last patient whose data was included in the estimation of the TD.

These simulated trials have been repeated 1000 times and the mean estimate of the *TD* is found along with the mean trial length for each procedure. The variability of the TD estimates is shown by the 95% reference range (2.5 and 97.5 percentiles of the estimated *TD*). The ratio of the reference range limits shows the precision of the estimates. The minimum and maximum values are also recorded to show the extreme values that the estimates may take.

The proportion of trials that produced an estimate within 30% of the true TD ($366mg/m^2$) is also displayed. This interval is (256.2, 475.8). This was conducted to show the proportion of times the trial estimated a dose that was a balance between being efficacious and toxic.

5.5.1 Generated by the Proportional Odds Model with log(dose)

Results from the LRDP1, LRDP3 and ICSDP when the data are generated from a PO model with log(dose) as a covariate.

Design	LRDP1			LRDP3			ICSDP		
Variable	TD_{20}	No. of Cohorts		$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate	371.9	16.92		360.0	14.87		381.5	14.67	
(2.5, 97.5) percentiles of estimates	(227.9, 554.1)			(232.2, 538.8)			(247.7, 575.3)		
97.5/2.5	**2.43**			**2.32**			**2.32**		
Min	169.4	8		1.3	1		178.4	8	
Max	750.3	20		714.1	20		742.2	20	
% in (TD±30%)	82.2			85.3			83.0		
Precision	Safety	Max No.							
73.5	0.0	26.5		89.6	0.1	10.3	93.5	0.0	6.5

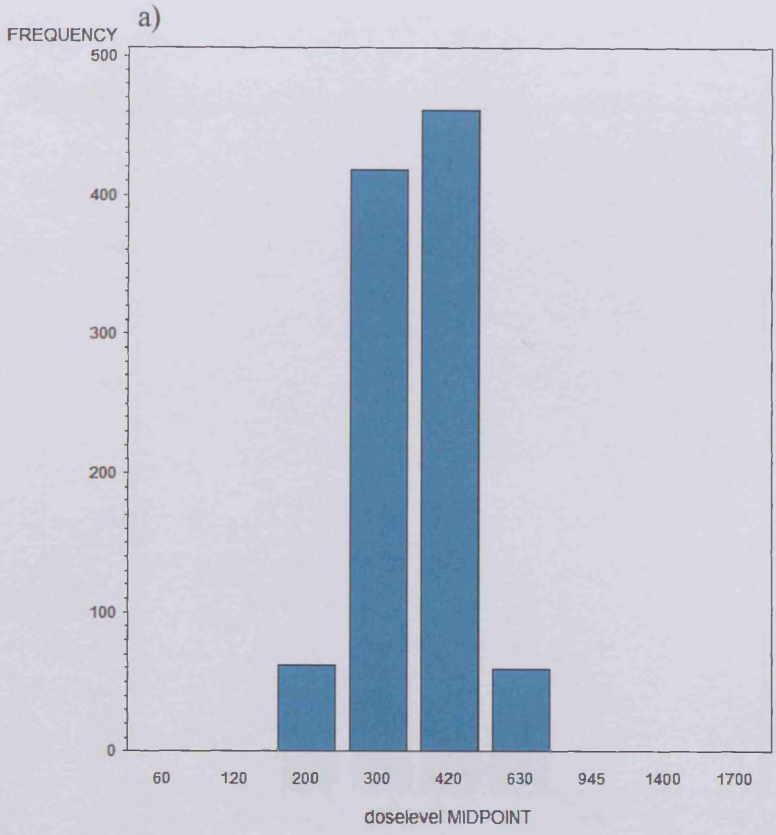
Table 5- 7: Results from 1000 trials, generated by the PO model and escalated with the patient gain. TD=$366mg/m^2$.

Table 5-7 shows that when the data are generated by the PO model with log(dose) as a covariate for 3 cycles, the best estimates of the TD are produced by the LRDP3, which is to be expected since the analysis models matches the generation model. The estimates produced by the ICSDP are slightly overestimated which is to be expected from Figure 5-2. The precision of the estimate is best for the LRDP3 and the ICSDP and the proportion of estimates within a 30% limit is highest for the LRDP3, but the ICSDP and LRDP1 also produce similar results. The precision associated with the estimates, which is shown by the ratio of the reference range limits, is similar for the LRDP3 and ICSDP but worst for the LRDP1 and the range of estimates is also larger for the LRDP1 than the ICSDP. The LRDP3 has an extreme minimum estimate (1.3) due to the occurrence of an event within cohort 1. Since no other information was known about the dose-response relationship then, the estimate of the TD produced was extremely low and the trial was stopped for safety. This actual event occurred in a later cycle which is why the other two procedures are not as affected by it. The LRDP1 did not observe this event and the ICSDP had other data on other doses to use with this occurrence so that it could be put into perspective.

The average number of cohorts was lowest for the ICSDP (14.67) due to the increased amount of information obtained. Although the average number of cohorts is quite low for the LRDP3, the actual average length of these trials would be $14.87 \times 3 = 44.61$ cycles due to having to wait for 3 cycles before starting a new cohort. The length of the trials when using the ICSDP and LRDP1 is equal to the number of cohorts included since each new cohort is recruited after every cycle. The LRDP1 required the most number of patients on average ($16.92 \times 3 = 50.76$) and the design which requires the fewest number of patients is the ICSDP ($14.67 \times 3 = 44.01$), but the LRDP3 is very similar ($14.87 \times 3 = 44.61$).

The simulated trials stop mostly due to the precision rule (particularly for the ICSDP) with most others stopping due to reaching the maximum number of cohorts. The LRDP1 produced the largest number of trials continuing until the maximum cohort had been recruited (20 cohorts). Furthermore, of the 73.5% achieving precision, approximately 7% achieve precision after recruiting the 20^{th} cohort, so actually over 34% of the simulated trials last for the maximum amount of time. The ICSDP also has some trials achieving precision in the final cohort, but only approximately 3%, so less than 10% of the trials last the maximum amount of time. For the LRDP3, approximately 90% achieve precision, with approximately 3% achieving it in cohort 20 so nearly 14% of trials last the maximum amount of time.

Although the analysis is done to compute a TD on the continuous scale, in reality the discrete dose levels would be recommended for further investigations. The distribution of these recommended dose levels are shown in Figure 5-7.



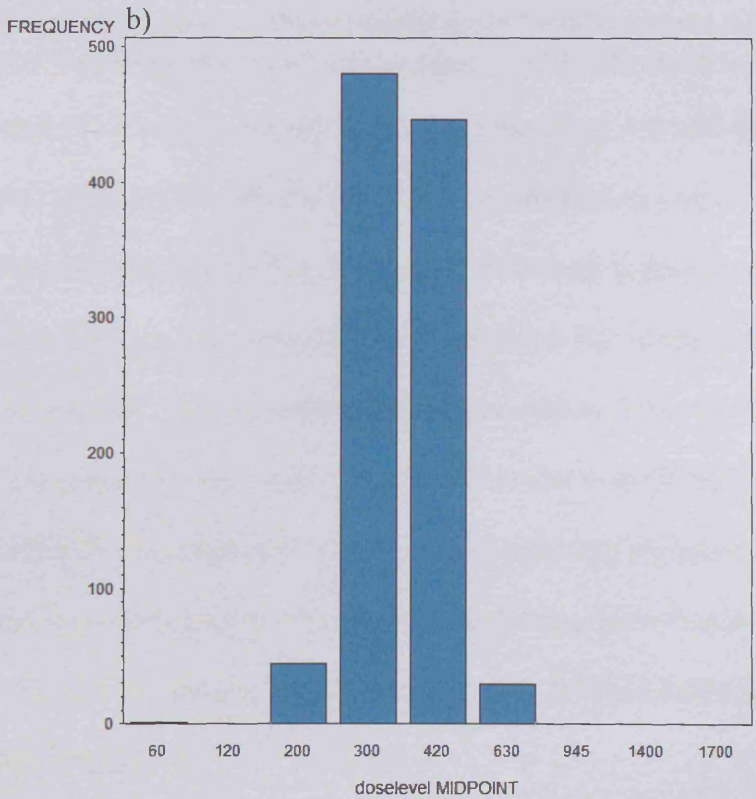


Figure 5- 7: Distribution of recommended doses for different procedures when the data is generated by the PO model with log(dose), a) LRDP1, b) LRDP3, c) ICSDP

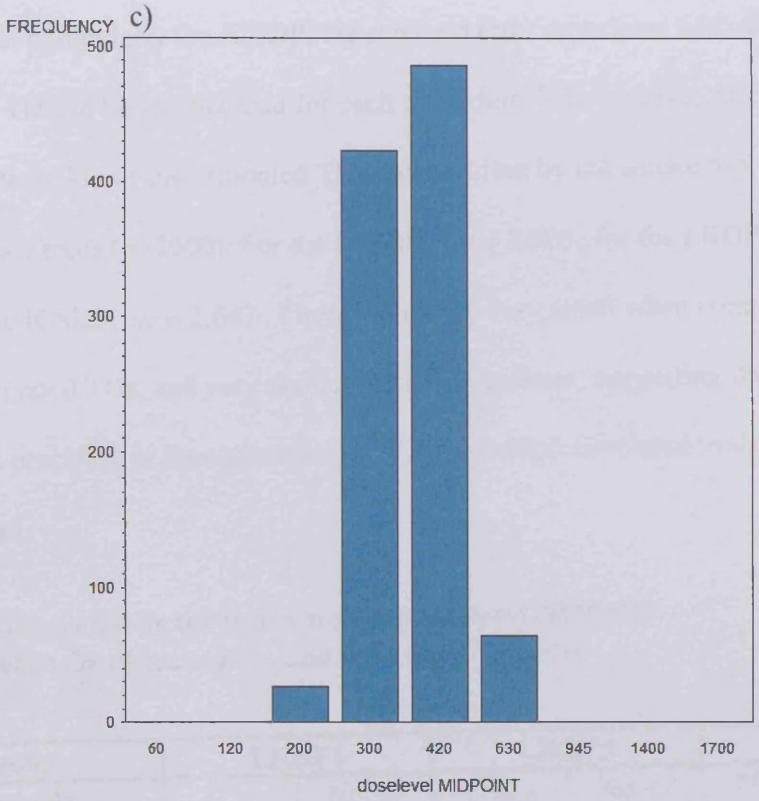


Figure 5-7 cont: Distribution of recommended doses for different procedures when the data is generated by the PO model with log(dose), a) LRDP1, b) LRDP3, c) ICSDP

Doses 300 and 420 are nearly equally distant from the TD of 366 with 420 being slightly closer. Figure 5-7 shows that the ICSDP has the highest proportion of dose levels recommended nearest the TD. The LRDP3 is the only procedure that recommends a dose that is not within 2 dose levels either side of the true TD, and that dose is the lowest dose. The proportion of trials that estimated the true TD within a 30% limit was greatest for the LRDP3, which was similar to the ICSDP. The LRDP1 estimated within this limit least (82% of the time), explaining the larger range of estimates and poorer consistency of estimates. This is clear from Figure 5-7 since the dose levels 200 and 630 are concluded more frequently for the LRDP1, which are outside of the 30% limit.

To ensure 1000 simulations is enough to obtain a good idea of the performance of the procedures, particularly the ICSDP, the standard error associated with the mean estimated TD can be summarized for each procedure. This involves calculating the standard deviation of the estimated TDs and dividing by the square root of the number of simulated trials (n=1000). For the LRDP1, $se = 2.805$, for the LRDP3, $se = 2.397$ and for the ICSDP, $se = 2.647$. These values are very small when compared to the mean estimated TDs, and very similar across procedures, suggesting that there is sufficient precision in the estimation of TDs from 1000 simulated trials for each of the procedures.

5.5.2 Generated by the Interval-Censored Survival Model

Results when the data are generated from the ICS model.

Design			LRDP1			LRDP3			ICSDP		
Variable			TD_{20}		No. of Cohorts	$TD_{31.6}$		No. of Cohorts	$TD_{31.6}$		No. of Cohorts
Mean estimate			371.3		17.13	354.1		14.80	371.2		15.02
(2.5, 97.5) percentiles of estimates			(217.7, 589.6)			(226.3, 523.6)			(243.2, 561.0)		
97.5/2.5			**2.71**			**2.31**			**2.31**		
Min			144.3		8	1.3		1	186.9		7
Max			994.6		20	683.7		20	690.4		20
% in (TD±30%)			78.5			83.4			85.5		
Precision	Safety	Max No.	90.6	3.6	5.8	89.3	0.6	10.1	90.4	0.0	9.6

Table 5- 8: Results from 1000 trials, generated by the ICS model and escalated with the patient gain. TD=366mg/m^2.

When data are generated from the ICS model, the LRDP1 and ICSDP produce almost identical estimates of the TD. However the precision of the ICSDP estimate is better and it is achieved on average over 2 cohorts earlier. This is down to the fact that the data generation model matches the analysis model. The LRDP1 produces similar estimates of the TD to the ICSDP since the ICS model generates data sequentially based on a 20% chance of DLT in cycle 1 which is required for the LRDP1. Although

there is some difference between the actual models (as shown in Figure 5-4), this difference is very small and simulation error could explain why the estimate is so good. The LRDP3 produces slightly worse estimates here. The precision of the TD estimates is again very good for the LRDP3 and the ICSDP, but the precision of the estimates produced by the LRDP1 is worse than when the data was generated by the PO model.

The average number of cohorts is largest for the LRDP1 (17.13) and has increased from when the data was simulated by the PO model. The average number of cohorts required for the LRDP3 is very similar to previously (14.80 vs. 14.67) and requires the fewest of the 3 procedures. The associated length of the trial would then be $14.80 \times 3 = 44.4$ cycles compared to the other procedures which require one cycle per cohort. The ICSDP therefore takes the shortest amount of time (15.02 cycles) and requires fewer patients than the LRDP1 ($15.02 \times 3 = 45.06$ vs. $17.13 \times 3 = 51.39$).

The LRDP3 stops for precision the majority of the time but does stop for safety in a very few cases. The trials stop for precision most often for the LRDP1 however the ICSDP stops for precision almost the same amount (90.6% vs. 90.4%). However, nearly 30% of the 90.6% stopping for precision with the LRDP1 achieve precision after cohort 20 has been recruited. This suggests that over 35% of the trials actually last the maximum trial length. This is in contrast to the ICSDP, where of the 90.4% stopping for precision, only 3.1% achieve precision after cohort 20. Just over 12% therefore require lasting the maximum amount of time. The LRDP3 has very consistent results to when the data was simulated by the PO model with regards to stopping for precision or reaching the maximum cohort. The LRDP1 now has to stop for safety reasons nearly 4% of the time, the LRDP3 less than 1% and the ICSDP not at all.

The distribution of recommended dose levels at the end of each trial is shown in Figure 5-8.

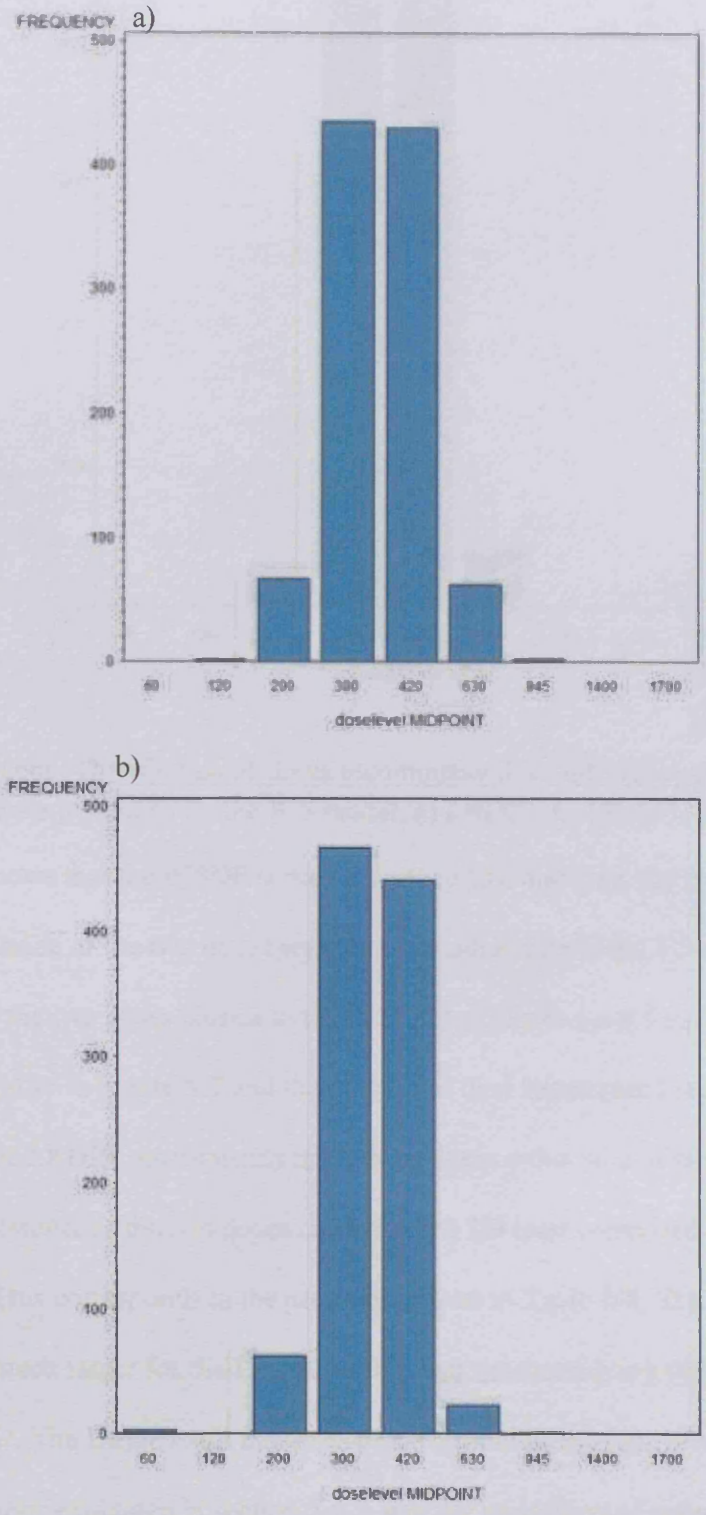


Figure 5- 8: Distribution of doses recommended for different procedures when the data is generated by the ICS model, a) LRDP1, b) LRDP3, c) ICSDP.

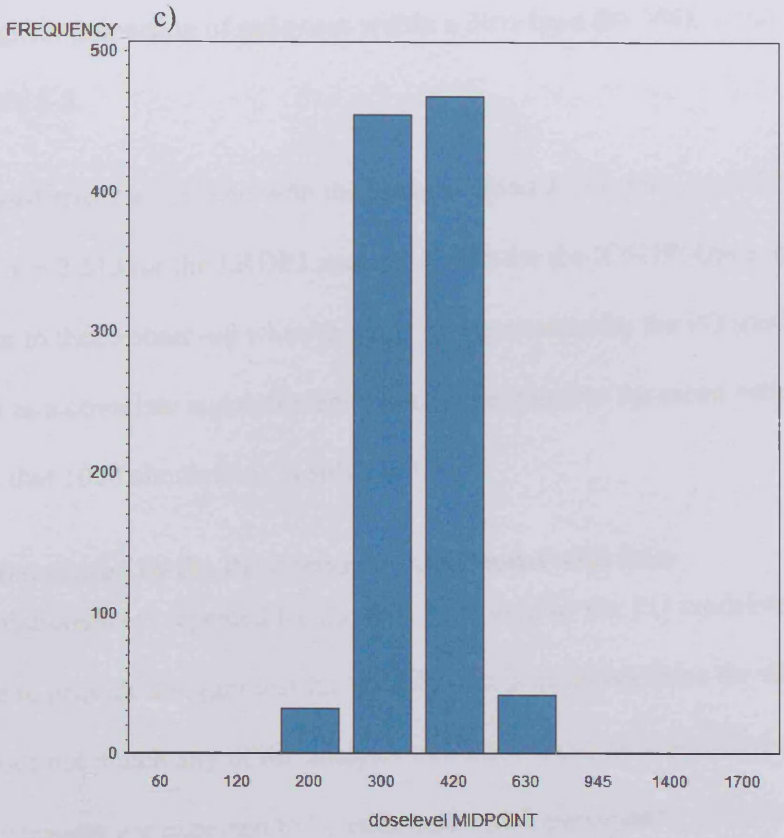


Figure 5-8 cont.: Distribution of doses recommended for different procedures when the data is generated by the ICS model, a) LRDP1, b) LRDP3, c) ICSDP.

Figure 5-8 shows that the ICSDP is the only procedure that does not recommend a dose level outside of the two dose levels that are either side of the TD and recommends the two doses closest to the TD (300, 420) the most frequently. The LRDP3 is similar to Figure 5-7 and the only other dose recommended is the lowest dose (60). The LRDP1 recommends up to three doses either side of the true TD and therefore recommends the two doses closest to the TD least compared to the other procedures. This corresponds to the precision shown in Table 5-8. The range of the estimates is much larger for the LRDP1 with fewer estimates lying within a 30% limit of the true TD. The LRDP3 still has an extreme minimum estimate of the TD due to the same reasons explained in section 5.5.1, and the proportion of estimates in the 30% limit of the TD is reduced now for the LRDP3 (83.4% vs. 85.3%). The ICSDP

has the highest proportion of estimates within a 30% limit (85.5%), which can be seen from Figure 5-8.

The standard errors associated with the mean estimated TDs are $se = 3.050$ for the LRDP1, $se = 2.513$ for the LRDP3 and $se = 2.525$ for the ICSDP. Once again these are similar to those observed when the data was generated by the PO model with log(dose) as a covariate and sufficiently small compared to the mean estimated TDs to conclude that 1000 simulations is sufficient.

5.5.3 Generated by the Proportional Odds Model with dose

The simulations were repeated for the data generated by the PO model with dose as a covariate to provide a bigger test for the different procedures since the data generation model does not match any of the analysis models. As can be seen in Figures 5-5 and 5-6, all estimates are expected to be underestimated, particularly those from the LRDP3.

The results are shown in Table 5-9.

Design			**LRDP1**			**LRDP3**			**ICSDP**		
Variable			TD_{20}		No. of Cohorts	$TD_{31.6}$		No. of Cohorts	$TD_{31.6}$		No. of Cohorts
Mean estimate			340.4		17.61	321.1		16.00	347.1		17.01
(2.5, 97.5) percentiles of estimates			(152.5, 553.2)			(1.3, 562.0)			(154.9, 576.6)		
97.5/2.5			**3.63**			**419.4**			**3.72**		
Min			52.6		3	1.3		1	6.6		3
Max			750.4		20	1250.4		20	1062.7		20
% in (TD±30%)			66.0			61.7			68.0		
Precision	Safety	Max No.	64.1	17.3	18.7	60.3	6.0	33.7	62.4	0.1	37.6

Table 5- 9: Results from 1000 trials, simulated by the PO model with dose and escalated with the patient gain. TD=366mg/m^2.

The results show that the ICSDP produces the best estimate of the TD although the precision is slightly better for the estimates produced by the LRDP1. The results agree

with Figures 5-5 and 5-6, with the best estimates being produced by the ICSDP although they are still underestimated. The LRDP3 produces much worse estimates.

The average number of cohorts required by the ICSDP is slightly higher compared to results in Table 5-7 and 5-8 (17.01 vs. 14,67, 15.02), but is still slightly less than that for the LRDP1 (17.13). The ICSDP therefore leads to the shortest trial length, despite the LRDP3 requiring fewer cohorts (14.8 cohorts).

The precision associated with the estimates is best for the LRDP1 with a similar result for the ICSDP, but is very poor for the LRDP3. The range of estimates for the LRDP3 (1.34, 1250.36) is extremely large, explaining the lack of precision.

The LRDP1 stopped due to safety reasons over 17% of the time. Figure 5-6 shows that the slope of the curve for the LRDP1 is much steeper than that for the true curve (PO with dose as a covariate). This suggests that lower doses correspond to a higher probability of DLT so the dose that is recommended by the model would cause less DLTs than expected (since the generation model has a lower P(DLT) for these doses). This would then cause the escalation to continue with an even higher dose level which might then be too toxic and cause the procedure to stop. Otherwise, the LRDP1 stopped for precision 64.1% of the time, with over 27% of those achieving precision once cohort 20 had been administered, so over 45% of the trials reached the maximum trial length allowed. The LRDP3 also stops for safety reasons an increased amount (over 6%) and only 60.28% stop for precision. Of those 60.28% an extra 4% achieve precision once cohort 20 has been recruited so just under 40% of trials last the maximum length. The ICSDP stops for safety 0.1% of the time and stops for precision over 62% of the time. Nearly 6% of the 62% achieve precision after cohort 20 has begun so over 43% of the trials last the maximum amount of time. Figure 5-5 shows

that the ICS curve is not as steep near the true TD as the LR curves. Combined with the addition of extra information at each cycle, this stops the procedure from recommending doses that are too high even though the true dose-response relationship will produce slightly less DLTs at the recommended doses than expected.

The distribution of recommended dose levels is shown in Figure 5-9.

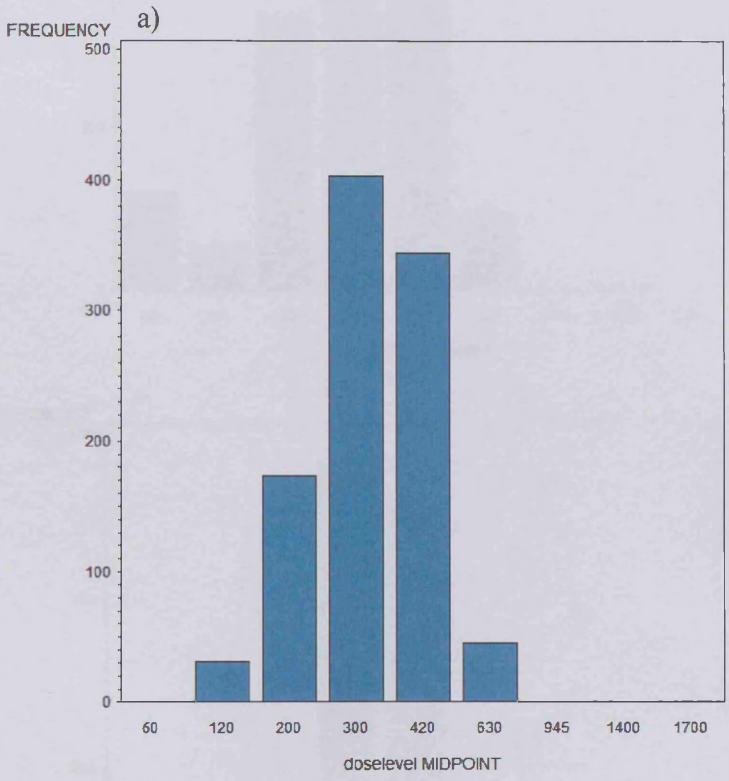
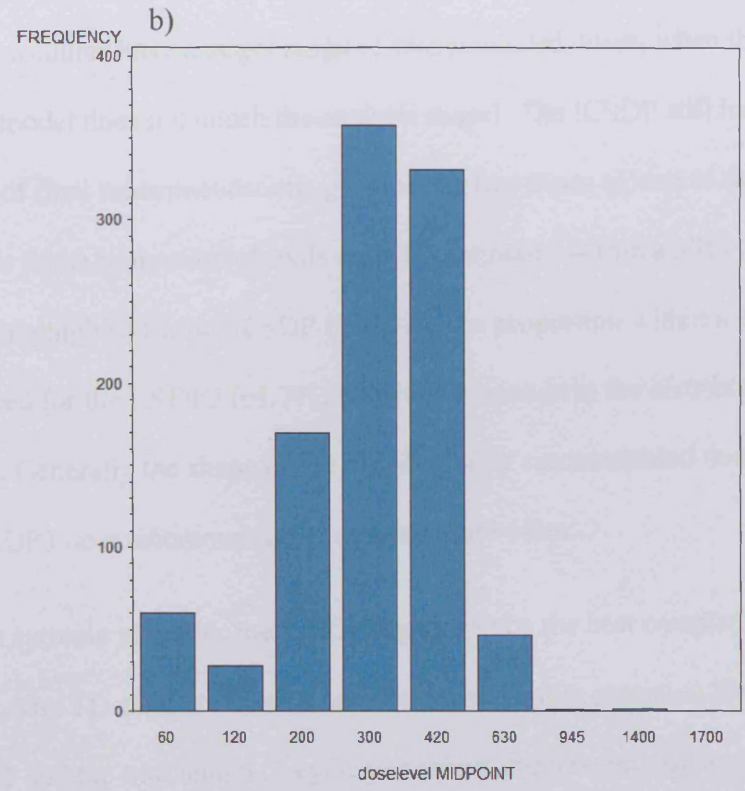


Figure 5- 9: Distribution of recommended doses for different procedures when the data is generated by the PO model with dose, a) LRDP1, b) LRDP3, c) ICSDP.



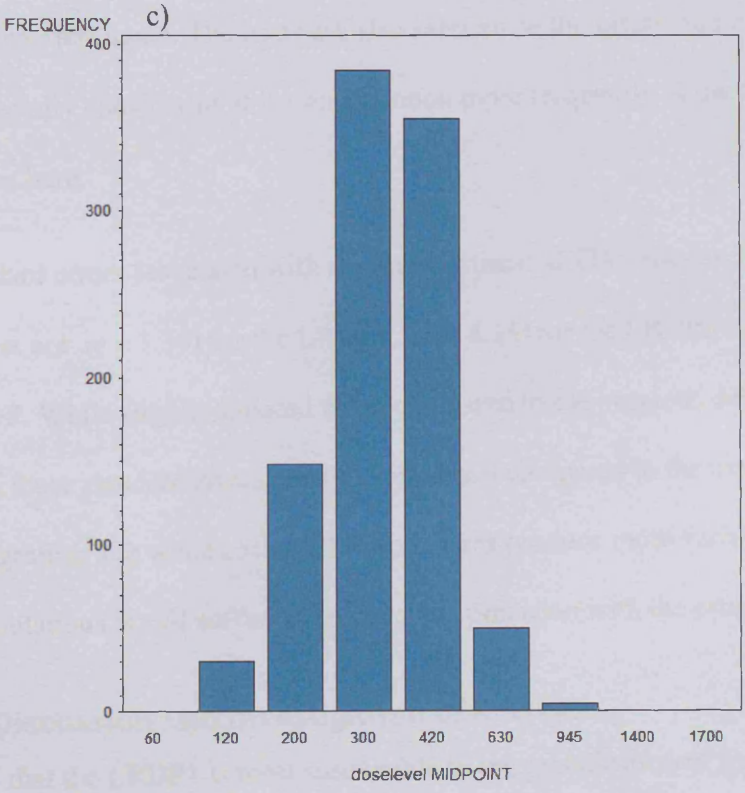


Figure 5-9 cont.: Distribution of recommended doses for different procedures when the data is generated by the PO model with dose, a) LRDP1, b) LRDP3, c) ICSDP.

All three procedures have a larger range of recommended doses, when the data generation model does not match the analysis model. The ICSDP still has the highest proportion of final recommendations given to the two doses closest to the TD which corresponds to the proportion of trials with TD estimates within a 30% limit of the true TD being highest for the ICSDP (68.0%). The proportion within a 30% limit is most reduced for the LRDP3 (61.7%) which corresponds to the distribution shown in Figure 5-9. Generally the shape of the distribution of recommended doses is similar, but the LRDP3 now recommends the dose 60 more often.

Under this extreme scenario, the ICSDP appears to be the best compromise for a procedure. The TD estimate is most accurate, with similar precision to that seen with the LRDP1 and the trial length (1 cycle per cohort implies the trial would last 17.01 cycles) is shortest again. The trials are also safer since the safety rule is hardly ever used, especially considering it is used so much more frequently in the other two procedures here.

The standard errors associated with the mean estimated TDs from each of the procedures are $se = 3.361$ for the LRDP1, $se = 4.351$ for the LRDP3 and $se = 3.564$ for the ICSDP. While slightly inflated when compared to the previous data generation methods, these standard errors are still very small compared to the mean estimated TDs suggesting that while each of the procedures produce more variable results here, 1000 simulations is still sufficient to conclude precision with the estimated TDs.

5.6 Discussion and Investigation of Results

It seems that the LRDP1 is most susceptible to misspecification of the analysis model to the data generation model. Although the TD estimates remain quite consistent, the precision associated with them becomes worse, as does the expected number of

cohorts required and the proportion of times the estimate is within a 30% limit of the TD. The reasons for stopping also change most for the LRDP1, with an increasing proportion of trials stopping for safety reasons as the data generation model deviates increasingly from the analysis model. The LRDP3 and the ICSDP produce quite consistent results with very slight improvements when the analysis model matches the data generation model. The main benefit of the ICSDP over the LRDP3 is in the actual length of time required to run the trial. Despite needing slightly fewer cohorts to achieve accurate and consistent estimates, the LRDP3 requires much longer due to the nature of having to wait for each patient to complete 3 cycles of therapy before enrolling the next cohort. This also creates an issue with some early cohorts since if an event happens in cohort 1 after cycle 1, no other information allows the procedure to continue and the drug is deemed unsafe. This does not happen with the ICSDP because the constant accrual of information from every cycle ensures that more information is obtained at once to allow perspective for the events that might occur (due to the random nature of simulation) on the very low doses.

5.7 Continuing the procedures to the maximum number of cohorts

Investigating the use of the procedures when the maximal amount of information is obtained may provide further insight.

Design	LRDP1		LRDP3		ICSDP	
Variable	TD_{20}	No. of Cohorts	$TD_{31.6}$	No. of Cohorts	$TD_{31.6}$	No. of Cohorts
Mean estimate	367.3	20	360.0	19.94	362.1	20
(2.5, 97.5) percentiles of estimates **97.5/2.5**	(216.7, 558.6) **2.58**		(238.3, 506.6) **2.13**		(218.8, 536.8) **2.45**	
Min	144.3	20	1.34	1	144.9	20
Max	994.5	20	628.1	20	777.1	20
% in (TD±30%)	82.3		89.7		80.0	

Table 5- 10: Results from 1000 simulations by the ICS model with the patient gain and a fixed number of patients (20 cohorts of size 3). TD=366mg/m^2.

Table 5-10 shows the results from continuing all procedures until 20 cohorts of patients have completed the trial. All of the mean estimates of the TD are closer to the true TD of 366 than in Table 5-8. Compared with the LRDP1 the ICSDP has greater precision, there are slightly fewer estimates within a 30% limit of the true TD, but the range of the estimates is not as extreme. The LRDP3 still has one occurrence where the trial stopped for safety after observing one cohort. The ICSDP performs slightly worse than in Table 5-8, which is due to the asymptotic variance for the ICS model. This variance relies on information from the number of patients administered to a given dose level as well as the number of toxicities observed at that dose. After the precision of the TD estimate satisfies the precision stopping criterion, if the escalation continues, more doses may be administered. When these doses are ones that have been slightly underrepresented previously, the inclusion of these additional data can cause the asymptotic variance to increase once again. In many cases, the ratio of the credible interval may then creep back above the threshold for which precision is claimed, even if the precision stopping criterion has previously been met. In this case, the final

estimate of the TD is no longer deemed as precise as if the trial stopped when precision was first achieved, and the variability across the trials is therefore increased.

5.8 Considering the effect of censoring

The effect of censored data on the performance of the dose escalation procedures is an important consideration. A natural benefit of the ICS model is its ability to limit the effect of non-informative censoring due to the fact that only the cycles of information after the occurrence of censoring are lost. The cycles of therapy prior to the occurrence of censoring are still analysed and contribute to the estimation of the TD. One can incorporate non-informative censoring into the simulated datasets to investigate how the different procedures react.

Furthermore, it may also be appropriate to consider the effect of informative censoring. Informative censoring could arise due to the fact that patients who are likely to experience a DLT may be more likely to withdraw prior to experiencing this DLT due to intolerance to the investigational drug. This intolerance could manifest itself in the way of an increased occurrence of lower grade toxicities (LGTs) in cycles prior to the cycle that a DLT may actually occur in. Patients who experience a large number of LGTs in early cycles may be more likely to cease treatment after these cycles and therefore would not contribute a DLT from a later cycle to the analysis, even though they may be more likely to experience one.

Non-informative censoring is easy to implement through the use of a Bernoulli random variable. If it can be assumed that all patients have a 10% chance of being censored in each cycle, a Bernoulli random variable can be simulated with a probability of 0.1 for each patient and each cycle. If a patient is censored in cycle 1, cycles 2 and 3 will also be missing, however if a patient is censored in a later cycle,

earlier cycles will still contribute, so there will be a lower proportion of censored observations than of censored patients.

The expected magnitude of censoring for the ICSDP is displayed in Table 5-11.

	10% censoring in cycle 1.	10% censoring in cycle 2.	10% censoring in cycle 3.	Total loss
# starting	$n_1=60$	$n_2=43$	$n_3=34$	
#censored	$n_1^c=6$	$n_2^c=4.3=>5$	$n_3^c=3.4=>4$	15 patients
Total #cycles lost to follow up	$3n_1^c=18$	$2n_2^c=10$	$n_3^c=4$	36 cycles
E(#DLTs) at TD	$n_1^{DLT}=12$	$n_2^{DLT}=4.3$	$n_3^{DLT}=1.7$	
#DLTs censored, observed #DLTs	$n_1^{DLT,c}=1.2$ $n_1^{DLT}=12-1.2$ $=10.8=>11$	$n_2^{DLT,c}=0.43$ $n_2^{DLT}=4.3-0.43$ $=3.87=>4$	$n_3^{DLT,c}=0.17$ $N_3^{DLT}=1.7-0.17$ $=1.53=>2$	1.8 DLTs
Observed P(DLT) at TD	10.8/60=18%	3.87/43=9%	1.53/34=4.5%	

Table 5- 11: Effect of non-informative censoring

A total of 15 patients would be expected to be censored throughout the trial which would also be true for the LRDP3. This is 25% of the patients who started treatment. For the LRDP1, clearly just 10% are censored since only the patients in the first cycle are affected by censoring.

The maximum number of patient cycles one would expect to see at the target dose for the ICSDP is 60+48+43=151, when considering the patients who withdraw to the occurrence of DLTs. When 36 patient cycles are censored, this corresponds to approximately 21% of cycles being censored, which is slightly less than the number of patients. One would then expect the ICSDP to perform slightly better than the LRDP3 when considering non-informative censoring since if a patient is censored, all information is lost, therefore the number of missing observation periods is the same as the number of missing patients so would correspond to 25% of lost information. The LRDP1 should not be too affected, since only 10% of observations in the first period

will be missing (6/60), however since the ICSDP still contains more data from later cycles, one would expect the information obtained overall to be less. There will be an expected 115 patient cycles worth of information at the target dose for the ICSDP as opposed to 54 patient cycles for the LRDP1.

The effect on the observed conditional P(DLT) at the TD is also reduced by 10% for all cycles, suggesting that observations at the TD will only correspond to 18% chance of DLT in cycle 1, 9% in cycle 2 and 4.5% in cycle 3. This corresponds to an overall chance of DLT of 28.7%, which is slightly less than 10% smaller than 31.6%. P(DLT)=0.18 in cycle 1 is 10% less than the expected 0.2, therefore the estimation may be less biased for the procedures looking at longer periods of time.

For informative censoring, only the patients who experience DLTs should be censored. In order to get a comparable rate of censoring to non-informative censoring in order to compare, it should be considered that only (on average) 31.6% of patients will experience a DLT over 3 cycles. In order to observe a rate of 10% overall for each cycle, approximately 1/3 of patients who have DLTs should be censored. Since the patients who are censored come from the subset of patients who have events, there are less missing cycles expected. Only the single observation of the actual event in that cycle is now missing since the cycles after the occurrence of DLT are not observed in the first instance. So there is only one missing cycle per patient.

	30% censoring of DLTs in cycle 1.	30% censoring of DLTs cycle 2.	30% censoring of DLTs cycle 3.	Total Loss
# starting	N_1=60	N_2=43	N_3=34	
E(#DLTs) at TD	N_1^{DLT}=12	N_2^{DLT}=4.3	N_3^{DLT}=1.7	
#DLTs censored, observed #DLTs	$n_1^{DLT,c}$=3.6 N_1^{DLT}=12-3.6 =8.4	$n_1^{DLT,c}$=1.29 N_2^{DLT}=4.3-1.29 =3.01	$n_3^{DLT,c}$=0.51 N_3^{DLT}=1.7-0.51 =1.19	5.4 DLTs
Observed P(DLT) at TD	8.4/60=14%	3.87/43=7%	1.53/34=3.5%	

Table 5- 12: Effect on informative censoring.

A total of 5.4 events are censored here, rounded to 6. This implies 6 cycles of therapy are missing which is approximately 4% of the maximum number of cycles expected at the TD. The ICSDP will therefore only lose 4% of patient cycles worth of information. 6 events being censored corresponds to 6 patients being censored, 10% of the total number starting treatment in cycle 1. The LRDP3 therefore loses 10% of information which is higher than for the ICSDP. In the LRDP1, 3.6 (rounded to 4) out of the possible 60 observations will be missing, approximately 7% of patient cycles, again higher than that for the ICSDP.

The observed conditional P(DLT) at the TD is 30% lower for each cycle. This suggests that the estimated TD will be higher than required. This reduction in conditional probabilities corresponds to an overall P(DLT) equal to 22.8%, which is less than 30% lower than 31.6%. The ICSDP and LRDP3 should be less biased than the LRDP1.

5.8.1 Results for Non-Informative Censoring

Results from the three procedures incorporating 10% non-informative censoring.

Design			LRDP1			LRDP3			ICSDP		
Variable			TD_{20}		No. of Cohorts	$TD_{31.6}$		No. of Cohorts	$TD_{31.6}$		No. of Cohorts
Mean estimate			365.5		17.39	357.0		15.45	369.7		16.81
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(203.9, 592.6) **2.91**			(224.9, 536.4) **2.39**			(241.8, 556.6) **2.30**		
Min			3.6		1	1.3		1	190.0		8
Max			1199.1		20	836.		1	20		752.3
% in (TD±30%)			75.8			83.4			85.7		
Precision	Safety	Max No.	67.7	1.0	31.3	85.3	0.3	14.4	77.9	0.0	22.1

Table 5- 13: Results from 1000 trials, simulated by the ICS model and escalated with the patient gain with 10% non-informative censoring. TD=366mg/m^2.

Table 5-13 shows that the mean estimates for the TD are very similar to those in Table 5-8, suggesting that non-informative censoring is not affecting the analysis very much. The precision of the estimates is slightly worse for the LRDPs and also the proportion of estimates within a 30% limit of the true TD is also reduced. The trials are also stopping less frequently for precision for all procedures, which is to be expected due to a decreased amount of information. The average trial length is also slightly longer for all procedures with the difference between the LRDP1 and ICSDP slightly reduced.

5.8.2 Results for Informative Censoring

Results from the three procedures with 10% informative censoring.

Design	LRDP1			LRDP3			ICSDP		
Variable	TD_{20}		No. of Cohorts	$TD_{31.6}$		No. of Cohorts	$TD_{31.6}$		No. of Cohorts
Mean estimate	455.0		16.76	431.8		15.12	457.3		14.73
(2.5, 97.5) percentiles of estimates	(254.9, 730.1)			(267.8, 627.6)			(292.3, 675.4)		
97.5/2.5	**2.86**			**2.34**			**2.31**		
Min	3.63		1	1.34		1	213.4		7
Max	1225.9		20	946.0		10	1032.9		20
% in (TD±30%)	58.0			67.2			58.3		
Precision / Safety / Max No.	75.7	0.2	24.1	87.9	0.3	11.8	88.7	0	11.3

Table 5- 14: Results from 1000 trials, simulated by the ICS model and escalated with the patient gain with 10% informative censoring. TD=366mg/m^2.

The mean estimates of the TD are now much worse than observed in Table 5-7. They are a lot larger than they should be and this is due to the fact that approximately one third of all DLTs that should occur are not observed. This then decreases the hazard associated with the drug and a higher dose appears to be tolerated at the TTL. The proportion of estimated TDs within a 30% limit of the true TD is also reduced since the point estimates are generally much higher. These differences are consistent across procedures, suggesting that informative censoring is not more of a problem for any one procedure. A comparison of the actual procedures still maintains that the ICSDP produces comparable results with the LRDP1 but has greater precision and finishes in a much shorter period of time, with trials stopping for precision more frequently.

5.9 The effect of incorrect assumptions in the trial design

To investigate the ICSDP in a more challenging setting, and perhaps a more realistic scenario, some of the underlying assumptions can be tested. The main assumption for the data generation is that, at the true target dose, the probability of a DLT halves conditionally for successive cycles. Currently this is reflected in the procedure through

the implementation of the pseudo data and also through defining the TD as the TD31.6. 31.6% is calculated from combining $\pi_{1,366} = 0.2, \pi_{2,366} = 0.1$ and $\pi_{3,366} = 0.05$ which is a direct result of the halving property. In reality the data may occur with this property but it cannot be predicted.

A further investigation is to be conducted where the ICSDP looks for a TD corresponding to a different level of toxicity despite the data being generated with the target dose of 366mg/m^2 and the conditional probabilities halving over cycles. This data generation method implies that the TD of 366mg/m^2 corresponds to a TTL of 31.6%.

When the conditional probabilities are mis-specified in the design phase, the procedure will be initiated with incorrect assumptions and a dose that is not necessarily the true TD will be investigated.

Table 5-15 shows the unconditional probabilities in the data generation model at the TD and compares to some alternative values that may be assumed when designing the trial. The reason for the choice of values will be given in the section 5.9.1, 5.9.2.

	$p_{TD,1} = \pi_{TD,1}$	$p_{TD,2}, \pi_{TD,2}$	$p_{TD,3}, \pi_{TD,3}$	$p_{TD}(c_3) = TTL$
Data Generation	0.2	0.08, 0.1	0.036, 0.05	0.316
Trial Design 1	0.2	0.09, 0.1125	0.04, 0.0563	0.33
Trial Design 2	0.15	0.128, 0.15	0.108, 0.15	0.386

Table 5- 15: Possible probability differences

Table 5-15 shows that when mis-specifying the assumptions based on how the P(DLT) changes over cycles can dramatically change the target toxicity level investigated. This is reiterated when looking at the range within 30% of the true TD. For a TTL of 31.6, the TD=366 mg/m^2 and $30\% \pm TTD$=(256,476). For a TTL of

33%, the TD=380 mg/m^2 and the 30% range is (266,494). For a TTL of 38%, the TD=435 mg/m^2 and the 30% range is (304,566). While the 30% range for the TD38 still contains the TD31.6 (366mg/m^2), it now doesn't contain the lower quartile of the 30% range associated with the TD31.6, so doses that would still be clinically meaningful would not be if the wrong TTL were investigated.

5.9.1 Investigating a TD33

The first scenario is Table 5-15 shows the assumptions that lead to a study investigating a TD33. A 33% chance of DLT is a simple choice of probability to investigate, since in rule-based dose-finding studies it is often a dose corresponding to 1/3 chance of DLT that has traditionally been investigated. Based on the true data generation parameters the dose that corresponds to a TTL of 33% is 380. The ICSDP should be able to estimate doses from the entire dose-response relationship so the test will be whether the estimated TD corresponds to the true TD33 of 380.

The error in assumptions will have most effect on the estimation of the TD33 when setting up the pseudo-data. The TD33 can be split into unconditional probabilities of 20% for the first cycle, 9% for the second cycle and 4% in the third cycle. This then corresponds to the conditional probabilities of $\pi_1 = 0.2, \pi_2 = 0.1125$ and $\pi_3 = 0.0563$ and the pseudo data will be presented as shown in Table 5-16.

	Dose $d_{(j)}$	$\pi^0_{(j)l}$	$n^0_{(j)l}$	$t^0_{(j)l} = n^0_{(j)l}\pi^0_{(j)l}$
ICSDP ***TTL=0.33***	$d_{(1)}$, cycle 1	0.2	3	0.6
	$d_{(1)}$, cycle 2	0.1125	2.4	0.27
	$d_{(1)}$, cycle 3	0.0563	2.13	0.1199
	$d_{(k)}$, cycle 1	0.5	3	1.5
	$d_{(k)}$, cycle 2	0.3098	1.5	0.4647
	$d_{(k)}$, cycle 3	0.1647	1.0353	0.1705

Table 5- 16: Pseudo-data for a procedure looking for the TD33.

The results for this procedure are shown in Table 5-17. Here the clinically relevant range based on 30% of the true TD (TTD=380) is (266,494).

Design			ICSDP		
Variable			TD_{33}		No. of Cohorts
Mean estimate			395.5		14.18
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(276.4, 587.7) **2.13**		
Min			257.17		8
Max			802.2		20
% in (TD±30%)			88.0		
Precision	Safety	Max No.	93.0	0.0	7.0

Table 5- 17: 100 simulations investigating a TD33. TD=380mg/m^2.

The true TD33 is 380, and the estimate here after just 100 simulations is very good. The precision of this estimate is also very good. A very high proportion of trials stop for precision with a high proportion of trials estimating within the clinically relevant range.

5.9.2 Investigating a TD38.6

The true target dose of 380 and the TTL of 33% is very similar to that for the TD31.6 so an even larger deviation can be considered which corresponds to completely incorrect assumptions such as a constant conditional probability of DLT over cycles. The assumption of a non-decreasing chance of DLT over cycles may be implemented where $\pi_1 = 0.15 = \pi_2 = \pi_3$. As shown in Table 5-15, this corresponds to an overall TTL of 0.386. Based on the true data generation model, the dose corresponding to this TTL is 435mg/m^2. The pseudo data is then as shown in Table 5-18.

	Dose $d_{(j)}$	$\pi^0_{(j)l}$	$n^0_{(j)l}$	$t^0_{(j)l} = n^0_{(j)l}\pi^0_{(j)l}$
ICSDP ***TTL=0.386***	$d_{(1)}$, cycle 1	0.15	3	0.45
	$d_{(1)}$, cycle 2	0.15	2.55	0.3825
	$d_{(1)}$, cycle 3	0.15	2.1675	0.3251
	$d_{(k)}$, cycle 1	0.5	3	1.5
	$d_{(k)}$, cycle 2	0.5	1.5	0.75
	$d_{(k)}$, cycle 3	0.5	0.75	0.375

Table 5- 18: Pseudo-data for procedure looking for TD38.6.

The results for this procedure are shown in Table 5-19. Here the clinically relevant range based on 30% of the true TD (TTD=435) is (304,566).

Design			**ICSDP**		
Variable			$TD_{38.6}$	No. of Cohorts	
Mean estimate			432.8	11.34	
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(268.0, 608.3) **2.27**		
Min			20.2	2	
Max			792.6	20	
% in (TD±30%)			89.0		
Precision	Safety	Max No.	93.0	6.0	1.0

Table 5- 19: 100 simulations investigating a TD38.6. TD=435mg/m^2.

Compared to the true TD of 435, this estimate is again produced very well with just 100 simulations. The precision of the estimate and the proportion of estimates in the clinically relevant range is again very good. Although there are a high proportion of trials stopping for precision here, there is also a noticeable amount of trials stopping for safety. This is to be expected, since the safety stopping rule dictates that if the dose to be administered is >0.44, the trial should stop for safety. Compared to 31.6% and 33%, p>0.44 is quite far away from the TTL (>11%), however compared to this procedure it is only just over 5% away from the TTL. Therefore repeating the procedure with a slightly higher safety probability should be recommended. The

safety probability is set to 0.5 now since this is again just greater than 11% from the TTL of 38.6% so a more comparable procedure should be produced. These results are shown in Table 5-20.

Design			ICSDP		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			435.2		11.52
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(284.6, 608.3) **2.14**		
Min			252.7		7
Max			728.4		20
% in (TD±30%)			89.0		
Precision	Safety	Max No.	98.0	-	2.0

Table 5- 20: 100 simulations investigating a TD38.6 with safety stopping rule p=0.5. TD=435mg/m^2.

The estimates here are much improved with better precision and no occurrences of stopping for safety.

In all of these additional procedures, the expected number of cohorts required is less than for the TD31.6 and the estimates seem to be produced better with increased precision. This is to be expected with procedures analysing binary observations since a probability closer to 0.5 is closer to the expected value of the binary endpoint. All observations of 0 (no DLT) and 1 (DLT) therefore have more of an equal contribution to the estimation of a dose corresponding to the TTL, so there is a maximal use of the information in the analysis of the data.

5.10 Conclusions

The best estimate of the target dose is generally produced the LRDP1 or the ICSDP and are within 30% of the true TD most often for the ICSDP.

However, in all possible circumstances, the ICSDP requires fewer cohorts than the LRDP1 and therefore is a shorter procedure. The LRDP3 always requires the fewest number of patients but due to the nature of having to wait for each cohort to finish 3 cycles of therapy before escalating, these designs will always take the longest to conduct.

The trials stop for precision most often for the ICSDP. The occurrences of safety stopping occur increasingly for the LRDPs as the data generation method deviates further from the analysis method. This is particularly true for the LRDP1 whereas this hardly ever occurs for the ICSDP.

The ICSDP is the most robust design, as determined as a compromise of good estimation of the TD and few cohorts, and has shown consistent results which are largely invariant to misspecification of the data generation model. It combines the benefits of both of the other designs by using all cycles so requiring less patients, and also escalating after every 1 cycle so the length of these trials are generally shorter with the same number of cycles as cohorts. The estimates are usually very good, if not the best of the three designs, and are generally produced with the best precision. It seems therefore that the best design of the three is the escalation by the ICSDP.

The patient gain function is the most ethical gain function to use in the setting of dose-finding trials in cancer treatments. Further gain functions will be explored for the ICSDP in Chapter 6 to investigate whether using the patient gain due to ethics is detrimental to the estimation of the TD. Chapter 6 will also investigate the incorporation of intra-patient dose adjustments into the ICSDP.

The ICSDP can also produce good estimates of target doses corresponding to different TTLs even when the assumptions are mis-specified and implemented in the pseudo-data.

6. Simple variations on the ICSDP

Chapter 5 showed the efficiency and effectiveness of using the ICS model within a Bayesian Decision Procedure. The realistic use of the procedure in the cancer setting is to use the patient gain function since it is ethical to treat cancer patients (who are usually the subjects in dose-finding trials for cancer therapies), however there are other gain functions that could be considered. The first part of this chapter will investigate the use of some other gain functions, to see whether restricting the procedure to the use of the patient gain is actually detrimental to the overall procedure.

Furthermore, an attraction of the ICSDP is that analysing data at the end of every cycle not only allows dose escalations to proceed as fast as when only one cycle of therapy is observed, but may also allow the opportunity for patients to change doses between cycles. This should prevent extended exposure to under/overdosing, since doses can change to what is currently believed to be the TD rather than what was believed when less information was known. This may also allow quicker convergence to the final determined TD estimate since the dosing will become more targeted. The second part of this chapter investigates utilising intra-patient adjustments within the standard ICSDP with the patient gain to investigate whether estimation of the TD can be improved upon.

6.1 Investigating the use of the variance gain function

The variance gain function discussed by Whitehead and Brunier [9] and described in section 4.3 of this thesis is defined as:

$$g_{i(J)} = \left(\frac{1}{\mathrm{var}\left(\log\left(TD_{i,TTL}^{(+J)}\right)\right)} \right),$$

where $TD_{i,TTL}^{(+J)}$ is the estimate of the TD corresponding to the TTL after the i^{th} observation, incorporating a set of doses J that can be administered to the next cohort. This set can consist of different dose levels for each patient in a cohort, as a combination may reduce the variance the most. The combination that minimises the variance of the estimated log(TD) given the current parameter estimates, and therefore maximises this gain, is administered.

The variance term calculated for this gain function is the expected asymptotic variance of the estimate of $\log\left(TD_{TTL}^{(+J)}\right)$ after i observations. This involves using the delta method as described in Chapters 2.2.3 and 4.3.

$$\text{var}\left(\log\left(TD_{i,TTL}^{(+J)}\right)\right) = \nabla\left(\log\left(TD_{TTL}\right)\right)^T I_E^{-1}(\boldsymbol{\varphi})\nabla\left(\log\left(TD_{TTL}\right)\right), \qquad (6.1)$$

where I_E^{-1} is the Expected Information Matrix found from twice differentiating the log-likelihood with respect to each parameter and taking the expectation of each element of the matrix. $\boldsymbol{\varphi}$ is the vector of parameters used in each procedure. For the LRDPs it consists of α_1 or α_3 and β, for the ICSDP it is $\theta, \gamma_1, \gamma_2$ and γ_3. The derivation of the asymptotic variance for the LRDP is given in [14]. This expected asymptotic variance uses the expected observations for the set of doses J with the current estimates of the parameters, along with the observed observations so far in the likelihood, to determine which set of doses would reduce the variance the most. The set of doses that reduces the expected asymptotic variance the most is administered to the next cohort.

The same set of simulation studies as described in section 5.2-5.5 are repeated here with the variance gain function used for escalation rather than the patient gain function. Data is generated from the PO model, the ICS model and the PO model with dose as a covariate (rather than log(dose)).

The pseudo-data also remains the same as described in Chapter 5. Since the variance gain function does not necessarily lead to the lowest dose being administered to all patients in the first cohort, the first cohort is forced to receive the lowest dose for safety reasons, and the pseudo-data is subsequently used to control how quickly the escalation proceeds to the estimated TD.

The same stopping criteria are also implemented. The procedure will continue until either a precision rule is achieved (the ratio of the exponentiated asymptotic credible interval limits for the estimate of log(TD) is <R, R=4), a safety rule is breached (P(DLT)>0.44 for the procedures with 3 cycles, P(DLT)>0.3 for the LRDP1) for a dose selected to administer) or the maximum number of patients have been recruited (20 cohorts=60 patients). The asymptotic credible interval is that obtained from the observed asymptotic variance of log(TD). That is the observed asymptotic variance obtained by using the delta method (as in Chapter 4.4) with current estimates of the parameters and using just the observed data from the trial so far in the likelihood, including the pseudo-data.

Once the procedure has ceased due to one of the stopping criteria, the estimated TD is outputted along with the number of cohorts required to achieve it, and the stopping reason. For all of the simulated trials, the average estimated TD is produced along with a reference range displaying the 2.5^{th} and 97.5^{th} percentiles of the estimated TDs from all of the simulations, and the ratio of the 97.5^{th} percentile to the 2.5^{th} percentile is calculated to show the precision of the estimates. The proportion of trials producing an estimated TD within a clinically meaningful range of the True TD ($TTD \pm 30\%$) is then displayed along with the percentage of trials stopping for each criterion.

Since the use of the variance gain is not particularly ethical due to the focus on estimating the dose-response relationship rather than dosing patients at the believed target dose, and this investigation is being conducted for information purposes, only 100 simulations have been conducted. This should highlight any major differences between the procedures, but the precision of the estimates may not be quite as good.

6.1.1 Generation by the Proportional Odds model with log(dose)

Results for the LRDP1, LRDP3 and ICSDP when the data is generated by the Proportional Odds model with log(dose) as a covariate is shown in Table 6-1.

Design	**LRDP1**			**LRDP3**			**ICSDP**		
Variable	TD_{20}	No. of Cohorts		$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate	204.3	2.84		287.3	6.66		153.6	1	
(2.5, 97.5) percentiles of estimates	(142.8, 242.7)			(135.3, 469.1)			(24.4, 160.4)		
97.5/2.5	**1.70**			**3.47**			**6.58**		
Min	142.8	2		1.3	1		24.4	1	
Max	268.6	15		563.3	17		160.4	1	
% in (TD±30%)	2.0			50.0			0.0		
Precision / Safety / Max No.	1.0	99.0	0.0	28.0	72.0	0.0	0.0	100.0	0.0

Table 6- 1: Results from 100 trials, simulated by the PO model and escalated with the variance gain. TD=366mg/m^2.

Table 6-1 shows that when the data are generated by the PO model with log(dose), the TD estimates are very poor for all procedures, particularly for the ICSDP. The LRDP3 produces the best result and stops for precision 28% of the time, but stopped for safety in the remaining trials. The average number of cohorts is also extremely low with the majority, or all in the case of the ICSDP, of the trials stopping for safety after the first cohort has been observed since the doses selected for the second cohort were deemed to be unsafe. This explains why the mean trial length and the mean TD estimate are so small, as very few of the trials lasted long enough trial to obtain a sensible estimate since they stopped for safety very early. The precision measured by the ratio of the

97.5/2.5 percentiles is not particularly useful here since the estimates are so underestimated due to early stopping.

The asymptotic variance is reduced the most when high and low doses are administered. This explains why the majority of trials stopped for safety since very high doses would have been attempted to be administered. It can also be assumed that for trials that did not stop for safety, only very low doses were administered in order to obtain as much information about the model and produce good estimates.

In order to investigate this further, the same results were looked at with all the trials that were stopped for safety removed. The results are shown in Table 6-2.

Design	**LRDP1**		**LRDP3**		**ICSDP**	
Variable	TD_{20}	No. of Cohorts	$TD_{31.6}$	No. of Cohorts	$TD_{31.6}$	No. of Cohorts
n	1		28		0	
Mean estimate	213.9	15	353.3	11.43	-	-
(2.5, 97.5) percentiles of estimates **97.5/2.5**	-		(220.5, 501.2) **2.27**		-	
Min	213.9	15	220.5	8	-	-
Max	213.9	15	563.3	17	-	-
% in (TD±30%)	0.0		85.7		-	

Table 6- 2: Results from 100 trials simulated by the PO model and escalated with the variance gain without trials stopped for safety. TD=366mg/m^2.

The only interpretable results here are for the LRDP3. Here the estimated TD and its precision, along with the proportion of trials within a clinically meaningful range and the average number of cohorts is much better and very comparable to those obtained in Chapter 5, in Table 5-7. However there are still very few simulations contributing to this summary.

To make the use of the variance gain more appropriate it can be utilised in such a way that only doses that produce probabilities of toxicity (according to the current model) that are below the safety threshold are part of the set for the variance gain to choose from. The corresponding results from that simulation are shown in Table 6-3.

Design	LRDP1			LRDP3			ICSDP		
Variable	TD_{20}	No. of Cohorts		$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate	350.8	15.08		344.6	12.66		353.8	14.16	
(2.5, 97.5) percentiles of estimates	(224.4, 551.0)			(208.4, 581.9)			(192.2, 532.0)		
97.5/2.5	**2.46**			**2.79**			**2.77**		
Min	202.7	9		1.3	1		178.7	9	
Max	616.6	20		584.2	20		627.1	20	
% in (TD±30%)	81			82			87		
Precision / Safety / Max No.	89	0	11	98	1	1	98	0	2

Table 6- 3: Results from 100 trials, simulated by the PO model and escalated with the variance gain for permissible doses. TD=366mg/m^2.

All of the estimates are slightly underestimated when compared to the corresponding procedures with the patient gain, shown in chapter 5, Table 5-7. This can be put down to the fact that a restriction is imposed on the choice of doses but it is not necessarily the dose closest to the believed TD that is selected. The dose selected is from the restricted set but maximises the information, so is likely to be from the low end of the range. Little information about higher doses is therefore obtained so the estimation of the TD is skewed to the lower doses. The precision of all estimates is slightly worse also. The average number of cohorts required is slightly lower for all procedures and the proportion of trials stopping for precision is increased for all.

There is also an instance here where the safety rule is used in the LRDP3 despite the variance gain being only for permissible doses. In this case, an event occurred at the lowest dose in cycle 3 for a patient in cohort 1. This caused the estimated TD to be

very low and the estimated probability of a DLT was over the safety threshold for even the lowest dose. Therefore, no dose was permissible and the safety rule stopped the trial. This DLT did not affect the other procedures since it was not observed in the LRDP1, and the ICSDP had a lot more information obtained at the time of this event, so the impact was limited to one cycle for that patient, rather than all the information for that patient.

The distribution of doses recommended at the end of the trial is shown in Figure 6-1.

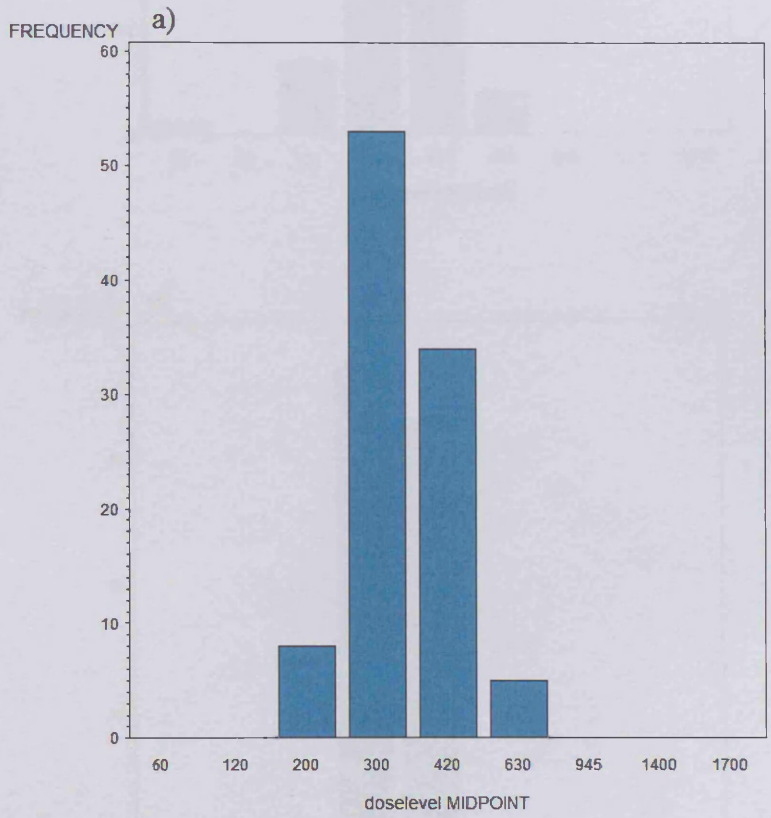
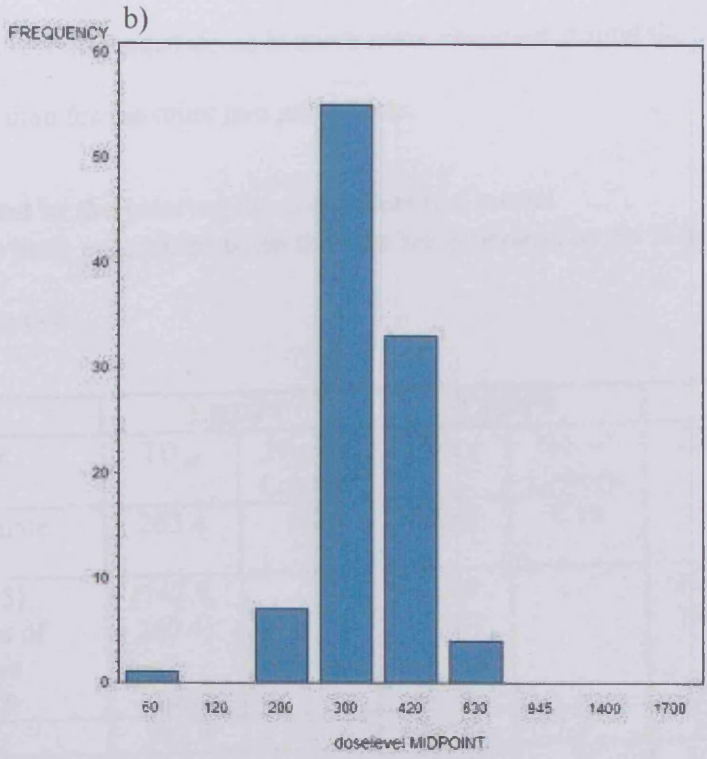


Figure 6- 1: Distribution of recommended doses for different procedures when the data is generated by the PO model with log(dose), a) LRDP1, b) LRDP3, c) ICSDP.



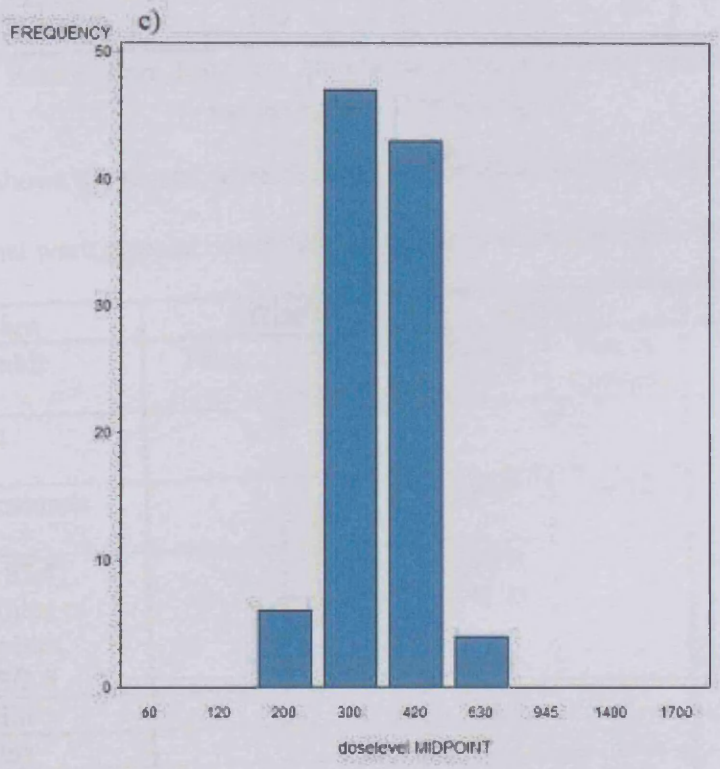


Figure 6-1 cont.: Distribution of recommended doses for different procedures when the data is generated by the PO model with log(dose), a) LRDP1, b) LRDP3, c) ICSDP.

The number of recommended doses is much more clustered around the true TD of 366 for the ICSDP than for the other two procedures.

6.1.2 Generated by the Interval-Censored Survival model

Results for the three procedures when the data are generated by the ICS model are shown in Table 6-4.

Design	LRDP1		LRDP3		ICSDP	
Variable	TD_{20}	No. of Cohorts	$TD_{31.6}$	No. of Cohorts	$TD_{31.6}$	No. of Cohorts
Mean estimate	203.4	3.2	271.6	6.16	155.0	1
(2.5, 97.5) percentiles of estimates	(142.8, 260.4)		(135.0, 479.4)		(24.4, 160.4)	
97.5/2.5	**1.82**		**3.56**		**6.58**	
Min	142.8	2	114.5	3	24.4	1
Max	268.8	13	509.9	14	160.4	1
% in (TD±30%)	4		36		0	
Precision / Safety / Max No.	0 / 100 / 0		25 / 75 / 0		0 / 100 / 0	

Table 6- 4: Results from 100 trials, simulated by the ICS model and escalated with the variance gain. TD=366mg/m^2.

Table 6-4 shows the results when the data is generated from the ICS model. Removing the trials that were stopped due to the safety rule is shown in Table 6-5.

Design	LRDP1		LRDP3		ICSDP	
Variable	TD_{20}	No. of Cohorts	$TD_{31.6}$	No. of Cohorts	$TD_{31.6}$	No. of Cohorts
n	0		25		0	
Mean estimate	-	-	294.7	10.52	155.0	1
(2.5, 97.5) percentiles of estimates	-		(201.9, 402.4)		(24.4, 160.4)	
97.5/2.5			**1.99**		**6.58**	
Min	-	-	201.9	8	24.4	1
Max	-	-	402.4	14	160.4	1
% in (TD±30%)	-		60.0		0	

Table 6- 5: Results from 100 trials, simulated by the ICS model and escalated with the variance gain without trials stopped by the safety rule.

Again the results are only interpretable for the LRDP3 which are more comparable to the patient gain, as shown in Chapter 5, Table 5-8. These are slightly worse than in Table 6-2 since there is the added discrepancy between the data generation and analysis models.

The results of investigating the variance gain but for permissible doses only are shown in Table 6-6.

Design	**LRDP1**			**LRDP3**			**ICSDP**		
Variable	TD_{20}	No. of Cohorts		$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate	346.2	15.11		336.5	11.79		346.4	13.30	
(2.5, 97.5) percentiles of estimates **97.5/2.5**	(181.8, 523.4) **2.88**			(203.6, 528.6) **2.60**			(214.8, 515.2) **2.40**		
Min	159.7	9		199.6	8		191.7	9	
Max	616.6	20		605.4	20		590.0	20	
% in (TD±30%)	75.0			67.0			87.0		
Precision	Safety	Max No.	84	0	16	99	0	1	98	0	2

Table 6- 6: Results from 100 trials, simulated by the ICS model and escalated with the variance gain for permissible doses. TD=366mg/m^2.

These results are once again comparable to those obtained in Table 6-3. All TD estimates are better than with the traditional variance gain but are underestimated when compared to the patient gain (Table 5-8) due to the restricted choice of doses for administration. Compared to Table 5-8, the precision is again worse. The average number of cohorts is less for all procedures with a higher proportion of trials stopping for precision, apart from for the LRDP1.

The distribution of the recommended doses is shown in Figure 6-2.

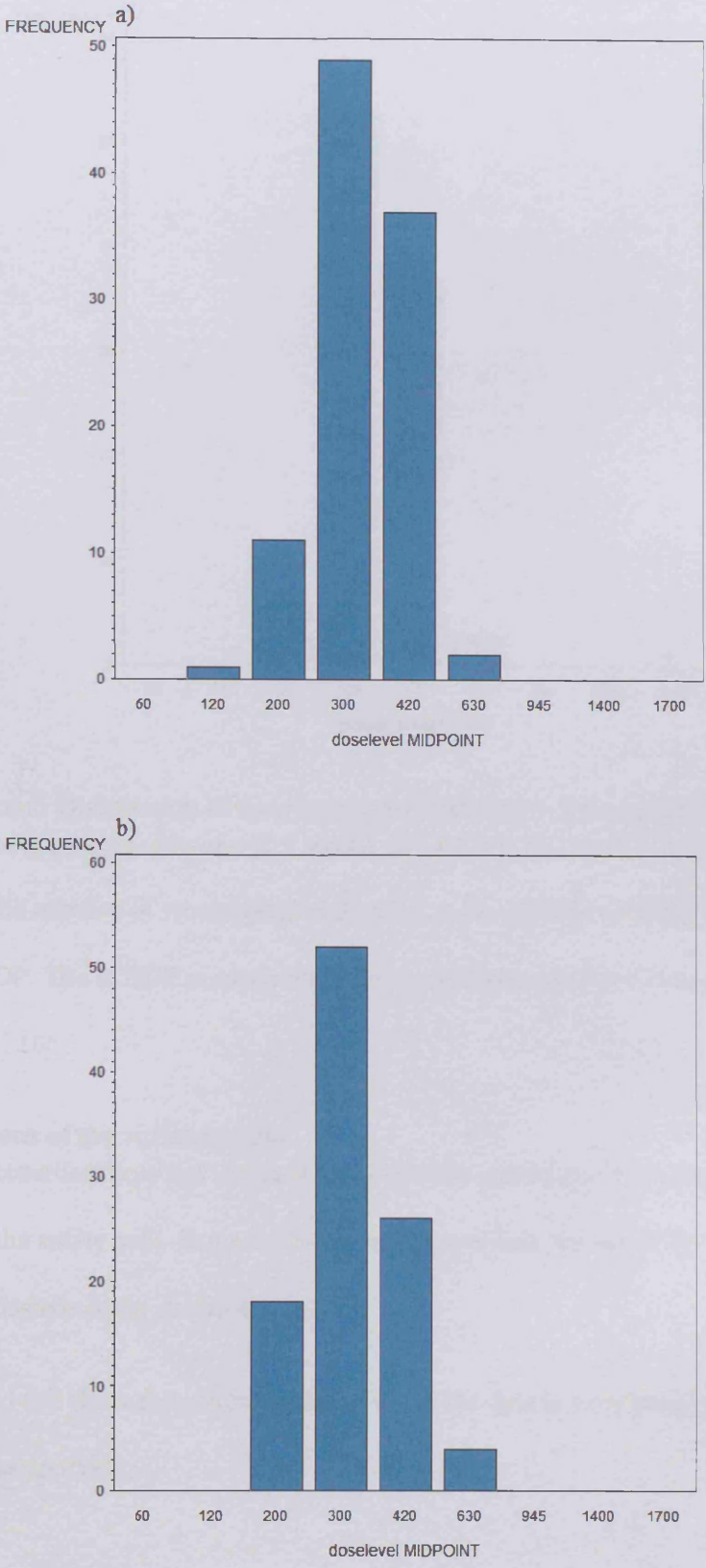


Figure 6- 2: Distribution of dose recommendations for different procedure when the data is generated by the ICS model, a) LRDP1, b) LRDP3, c) ICSDP.

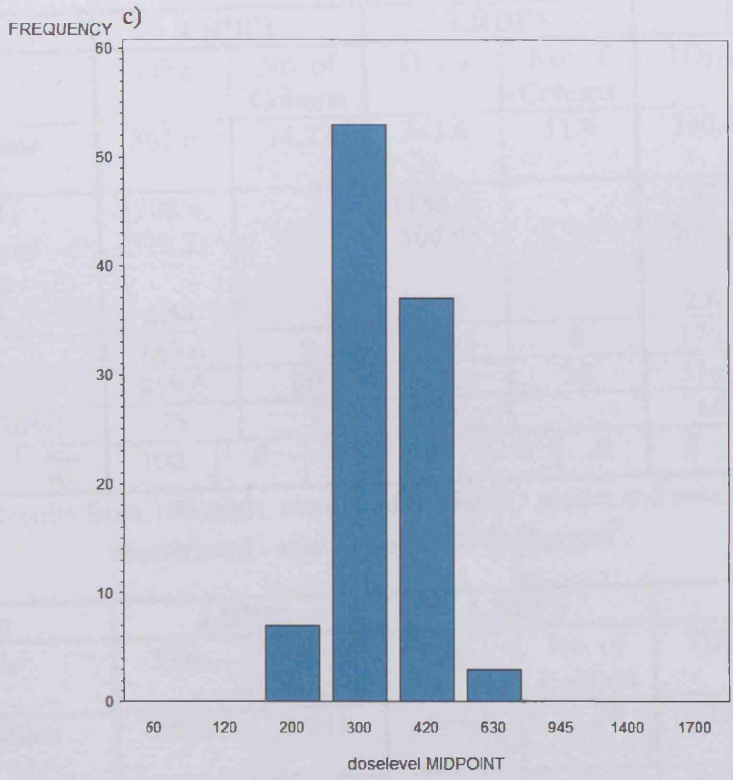


Figure 6-2 cont.: Distribution of dose recommendations for different procedure when the data is generated by the ICS model, a) LRDP1, b) LRDP3, c) ICSDP.

Once again, the number of recommended doses is more variable with the LRDPs than with the ICSDP. The ICSDP is much more clustered around the two doses closest to the true TD of 366.

6.1.3 Variations of the variance gain

This section considers how the variance gain function would perform when not restricted by the safety rule. Some further investigation into the use of the variance gain for permissible doses is also conducted.

Tables 6-7 and 6-8 show the results obtained when the data is simulated by the PO and ICS models respectively.

Design			LRDP1			LRDP3			ICSDP		
Variable			TD_{20}	No. of Cohorts		$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate			361.6	14.22		343.6	11.9		330.4	19.77	
(2.5, 97.5) percentiles of estimates			(208.4, 529.2)			(185.7, 506.0)			(192.1, 502.2)		
97.5/2.5			**2.54**			**2.73**			**2.61**		
Min			185.6	9		174.3	8		176.3	14	
Max			819.6	20		556.6	20		576.6	20	
% in (TD±30%)			75			79			66		
Precision	Safety	Max No.	100	0	0	100	0	0	8	0	92

Table 6- 7: Results from 100 trials, simulated by the PO model and escalated with the unrestricted variance gain. TD=366mg/m^2.

Design			LRDP1			LRDP3			ICSDP		
Variable			TD_{20}	No. of Cohorts		$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate			348.2	14.19		336.0	11.36		316.4	14.16	
(2.5, 97.5) percentiles of estimates			(200.1, 538.9)			(206.8, 511.5)			(181.1, 528.3)		
97.5/2.5			**2.69**			**2.47**			**2.92**		
Min			127.7	8		149.9	8		174.6	8	
Max			629.0	20		594.5	20		655.5	20	
% in (TD±30%)			77.5			80.3			53		
Precision	Safety	Max No.	100	0	0	100	0	0	98	0	2

Table 6- 8: Results from 100 trials, simulated by the ICS model and escalated with the unrestricted variance gain. TD=366mg/m^2.

The TD estimates produced by the unrestricted variance gain are clearly much better than when the variance gain was used with the safety stopping rule. The estimates are closer to the true TD with reasonable precision. The ICSDP does not perform as well here. The estimates are lower than the other procedures with fewest estimates within the 30% limit of the true TD. When the data is generated by the PO model, the required number of cohorts is very large for the ICSDP and the majority of trials stop for reaching the maximum number. This is largely down to the fact that the precision stopping rule is based on the asymptotic variance of the log(TD) estimate, which

depends very heavily on the data obtained so far for the ICS model, including the occurrence of DLTs, which the asymptotic variance for the PO model does not. When the data is generated from a different model, the observations do not necessarily occur in accordance with the predicted probability of the event. The variance includes the occurrence of toxicities and the estimated probability of DLT, when there is discordance between the two values, the variance does not reduce as quickly. When the data generation model does match, more data contributes to the variance calculation in accordance with the expected probabilities estimated by the model. The LRDP1 produces very reasonable results, albeit not as good as those obtained when using the patient gain (chapter 5, Table 5-7, 5-8).

The main issue with the unrestricted gain function is that it is unethical to allow a procedure to solely dictate which doses are administered to patients. In order to see why this escalation procedure is not ethical, one of the simulation trials can be considered in detail. Table 6-9 shows the dose recommendations for cohort 2 together with the probability of toxicity estimated after the observations of cohort 1 are obtained. Different doses are administered in some cases to the patients within the same cohort since the variance gain allows a combination of different dose levels to be administered if it should reduce the variance more.

	LRDP1, Cohort 2	**LRDP3, Cohort 2**	**ICSDP, Cohort 2**
Patient 4	d=1400, p=485	d=945, p=0.593	d=120, p=0.280
Patient 5	d=1400, p=0.485	d=945, p=0.593	d=630, p=0.526
Patient 6	d=1700, p=0.500	d=945, p=0.593	d=1700, p=0.704

Table 6- 9: Dose administrations recommended for cohort 2 when escalated by the unrestricted variance gain and simulated by the ICS model.

The number of times each dose is allocated as well as how early in the trial each dose is administered for the first time is shown in Table 6-10 as found from both LRDPs and the ICSDP escalations.

	Number of patients receiving dose		
	Earliest this dose is seen		
Dose	LRDP1, precision achieved after cohort 8	LRDP3, precision achieved after cohort 9	ICSDP, precision achieved after cohort 20
60	-	-	16
			Cohort 5
120	-	-	2
			Cohort 2
200	3	1	1
	Cohort 2	Cohort 2	Cohort 3
300	3	6	2
	Cohort 5	Cohort2	Cohort 5
420	6	8	7
	Cohort 3	Cohort 7	Cohort 7
630	6	3	9
	Cohort 3	Cohort 5	Cohort 2
945	-	3	1
		Cohort 3	Cohort 3
1400	2	3	15
	Cohort 4	Cohort 6	Cohort 6
1700	1	-	4
	Cohort 4		Cohort 2

Table 6- 10: Administration of doses in one simulated trial (escalated with the unrestricted variance gain and data generated by the ICS model) when the safety rule is not used, excluding cohort 1 and pseudo-data.

As expected, the variance gain is exploring very high doses very early on in the escalation procedure in order to obtain more information about the model. This is particularly evident with the ICSDP, where patient 6 (in cohort 2) is given the highest dose possible and some of the middle doses, are rarely explored, particularly 300. The reason this was more extreme in this case is because of the extra terms in the asymptotic variance function which depend directly on the number of toxic events and the probability of toxicity for different doses. Very high and low doses minimise the variance so these are the doses chosen most often (dose 60 is chosen 16 times and 1400, 15 times). This also explains the reasoning for the increased number of cohorts

required to complete the trial. Since fewer doses are chosen for exploration, more patients on the high and low doses are needed to produce the precision of results required.

With the LRDPs, one of the doses near the true TD is explored quite early, but the other is not until some time later in the escalation procedure (in the LRDP1 dose 420 is seen in cohort 3 but dose 300 is not seen until cohort 5, in the LRDP3 dose 300 is seen in cohort 2 but dose 420 is not seen until cohort 7). Clearly the administration of overly toxic doses so early on in the procedure is not ethical, particularly when the safety of the drug is not known at all and the purpose is to determine the TD safely.

So although the overall output from the trial is quite insightful, the trial is unethical. The trial also depends greatly on the model and few doses are explored around the true TD. This dependence would not be popular with clinicians as it eliminates the need for adapting doses to specific patient needs, particularly when a DLT might occur. The unrestricted variance gain function appears to work better for the LRDPs than the ICSDP, but the ethics are still unacceptable, regardless of what model is used.

The variance gain for permissible doses only, which allows the variance gain to only choose between doses that are deemed to be safe (as shown in Tables 6-3 and 6-6), provides much better results than using the variance gain alone (Tables 6-1 and 6-4). In reality, this is the way the variance gain would most likely be used since, as shown, it is neither safe nor ethical to allow a procedure to administer unsafe doses to extremely ill patients (since patients in dose-finding trials for cancer treatments are actual cancer patients). Although the TD estimates using the restricted variance gain (for permissible doses only) are better than those obtained by the original variance gain, they are not better than those obtained from using the patient gain (Tables 5-7, 5-

8) since they are further from the true TD and have greater variability. The required number of cohorts is generally slightly less however, and the proportion of trials stopping for precision is generally increased. One could maintain however that the idea of allocating doses that are not necessarily closest to the currently believed TD is still not ethical, despite the fact that all doses are now believed to not be toxic. In particular, as seen in the unrestricted variance gain, the procedure chooses doses at either end of the dose range more frequently to increase information about the dose response relationship. This is still true for the restricted dose range, so the dose levels near the target dose are still less likely to be administered than the lowest dose level and highest safe dose level.

One of the main attractions of using the variance gain is that the variance of the estimated TD should be reduced since the range of doses administered is larger. However this benefit is not particularly seen here. There could however be some issue with the way the precision is determined in each method (patient gain and variance gain for permissible doses) and whether it is valid to directly compare the two. With the variance gain, the precision (directly related to the asymptotic variance) is much better earlier on, after multiple administrations of one particular dose are observed and when doses at either end of the dose range are administered. The precision may then seem to be at the required level to cease the escalation, but the trial may actually not have had enough time to converge to a reasonable estimate and may also not have actually administered a dose near to the estimate it is producing. Also, in the restricted variance gain, there are much fewer doses for the procedure to choose between, so all contributions to the variance calculation are much more concentrated around a few doses and therefore reduction of the variance occurs much quicker.

In order to compare the results from the patient gain function and the variance gain function for permissible doses better, the same (average) number of cohorts that was observed when escalating with the patient gain for each procedure, is used as a stopping rule for the variance gain function for permissible doses instead of the precision rule. That is, when the trial reaches the same number of cohorts as was seen to be the average for the results by the patient gain function (as in Table 5-7), the trial will stop and estimate the TD then. Whether the ratio of the CI falls below a certain level (R<4) by the time the trial reaches the specified cohort will be recorded. The average estimate of the TD can then be computed and compared to the results of the patient gain. These results are shown in Table 6.11 where the data is simulated by both the PO model and the ICS model.

Design	**LRDP1**		**LRDP3**		**ICSDP**	
Variable	TD_{20}	No. of Cohorts	$TD_{31.6}$	No. of Cohorts	$TD_{31.6}$	No. of Cohorts
Mean estimate	342.4	17	346.0	15	342.5	15
(2.5, 97.5) percentiles of estimates	(220.3, 531.6)		(199.4, 508.3)		(164.1, 528.7)	
97.5/2.5	**2.41**		**2.55**		**3.22**	
Min	166.1		183.1		130.8	
Max	546.0		591.7		583.3	
% in (TD±30%)	84		91		77	
% with R<4	74		82		56	

Table 6- 11: Results from 100 trials, simulated by the PO model and escalated with the variance gain for permissible doses, until the same average number of cohorts as in the patient gain (Table 5-7) have been recruited.

Compared to Table 6-3, the LRDPs produce similar estimates of the TD but with better precision and a higher proportion of trials obtaining a TD within a 30% limit of the true TD. The proportion of trials achieving precision by this cohort is still quite high, although not as high as in Table 6-3. The ICSDP produces similar estimates here to the LRDP designs and also to the ICSDP in Table 6-3, but with worse precision and

the proportion of trials estimating a TD in a 30% limit of the true TD is smaller than in Table 6-3. The proportion of trials achieving precision at this cohort is less than that for the LRDPs, and also substantially smaller than the number of trials stopping for precision in Table 6-3.

Design	LRDP1		LRDP3		ICSDP	
Variable	TD_{20}	No. of Cohorts	$TD_{31.6}$	No. of Cohorts	$TD_{31.6}$	No. of Cohorts
Mean estimate	342.9	17	335.6	15	301.4	14.96
(2.5, 97.5) percentiles of estimates	(192.3, 493.6)		(209.7, 471.8)		(175.6, 486.0)	
97.5/2.5	**2.57**		**2.25**		**2.77**	
Min	181.1		184.5		170.2	11
Max	544.5		601.0		560.1	15
% in (TD±30%)	81		85		67	
% with R<4	71		88		72	
% safety stops	0		0		1	

Table 6- 12: Results from 100 trials, simulated by the ICS model and escalated with the variance gain for permissible doses until the same average number of cohorts as in the patient gain (Table 5-8) have been recruited.

When simulating by the ICS model, the results are similar to those in Table 6-11. The ICSDP is noticeably worse here when compared to Table 6-6. There is one occurrence here of a safety stop which explains why the estimate is quite obviously smaller. There is however, still a high proportion of trials achieving precision at the specified cohort for all procedures, however it is not as high as for the number of trials stopping for precision in Table 6-6. In each procedure, the average number of cohorts required to achieve precision for the variance gain with permissible doses (Table 6-6) was lower than the cohort this investigation forced the escalation to continue to, with a very high proportion of trials stopping for precision. By forcing the procedures to continue for longer allows the procedure to experiment with more doses, so the variance has the ability to increase again. The estimates are worse when forced to continue and precision is not concluded as frequently. When comparing these results

to those found when escalating with the patient gain alone (Table 5-8), the results are still not as good.

It seems then that the idea that the precision criterion is different for the patient and variance gain escalations may not actually be an issue. Although the variance gain seems to allow trials to stop earlier than the patient gain due to the direct link between the gain function and the precision criterion, forcing trials to continue for longer actually causes a detrimental effect to the estimates of the TD.

6.1.4 Conclusions

When using the simple variance gain, the LRDP3 produces the best estimates of the TD, which is better when the analysis model matches the data generation model. However, the variance gain produces very poor estimates and causes early termination of the trial very often due to trying to administer overly toxic doses, the trials that are not stopped for safety reasons generally administer very low doses. Administering very low doses in order to estimate the TD is not an ethical approach. Although toxic doses are not administered, administering sub-therapeutic doses is still not appropriate. Looking at an unrestricted variance gain increases understanding in this procedure but is unethical to use. The estimates are much better and more comparable to those obtained from the patient gain but with obviously fewer number of cohorts required to achieve them. Since the variance is derived from the likelihood function for the model, it depends on the expected probability of toxicity for each dose. The variance is therefore minimised when observations are obtained for doses with particularly low and high probabilities of DLT since knowledge of the dose-response relationship is required for the lower tail and therefore the upper tail too in order to understand the full model. Very high doses are therefore allocated very early on in the escalation procedure which is not at all ethical. Using a variance gain function for

permissible doses only does combat the problem of stopping early for safety and also produces more reasonable estimates of the TD, but these estimates are still not as good as those obtained by using the patient gain. The precision of the estimates is not particularly increased when using the variance gain which should be the main benefit of using the variance gain. The main advantage is the reduction in cohorts required and an increase in trials stopping for precision, however there is still the ethical consideration that the doses administered aren't the closest to the current estimated TD.

When comparing the use of the precision criterion for the different gain functions, it appears that the use of the variance gain does encourage trials to stop earlier. If these trials are forced to continue for longer, the estimates are not as good, as the procedure then begins to experiment with other doses in order to increase knowledge of the model so the estimates of the TD become more variable.

The variance gain, and its variations, seems to perform better for the LRDPs so for the ICSDP, it can be concluded to use the patient gain.

6.2 Incorporating intra-patient adjustments

The results from Chapter 5 show that incorporating as much information as possible from later cycles of therapy aids the escalation procedure, with shortened trials while maintaining the same level of precision of the estimated TDs. The patient gain is the most ethical gain function to use in the setting of Phase I escalation trials where the main interest is the safety profile of the drug. Furthermore, in section 6.1, the expected benefits of using another gain function such as the variance gain (the benefits being better precision of estimates) were unfounded and the allocation of doses that were not currently believed to be closest to the true TD suggested that other variations on the variance gain functions were not reasonable for use in the Phase I setting.

Building upon these results, further investigation of the ICSDP is now conducted in order to produce more ethical designs. Such designs may allow dose changes between cycles, dependant on the occurrence or not of DLTs. In particular, results from Chapter 5 suggest that more information, obtained sooner, produces a better escalation procedure. A natural extension then is to incorporate intra-patient adjustments (escalation/de-escalation), allowing the administered dose to be changed between cycles in order to allocate a dose that is deemed most ethical to administer at the current time. Rather than just starting new cohorts on the currently believed TD, all patients still in the trial will change their dose between cycles to also be on the currently believed TD. Theoretically, this should cause more patients to be treated at the currently believed TD and would avoid excessive over/under-dosing. The main issue with all patients being treated at the same dose is that the knowledge of the model across the dose range may be more limited than previously, so one may expect a trial to last longer than seen previously since it may take longer to achieve the precision required to cease the trial. Despite this expected increase in the length of the trial, the ethical nature of allowing patients to be treated at doses that are currently believed to be the target doses is something that should be considered.

An intra-patient escalation procedure will be conducted using the ICSDP with the patient gain and the results compared to the equivalent simulated trials from Chapter 5.

6.2.1 The ICSDP incorporating adjustments

To allow for the changing of doses between cycles, the existing simulated datasets as generated by a PO model and also an ICS model in Chapter 5 were used. This was done to again check the robustness of escalating according to an ICSDP. The PO model with dose as a covariate is not used here since the extreme test of the ICSDP

has been conducted in Chapter 5, so these comparisons are simply to provide slightly more insight into the new procedure.

The pseudo-data is of the same form as in chapter 5 and the stopping criteria remain the same. The precision rule, based on the ratio of the exponentiated asymptotic credible interval limits of the estimate of log(TD) ($R<4$), the safety rule, $P(DLT)>0.44$ for the dose chosen to administer, or the maximum number of cohorts are recruited, 20 cohorts of 3 patients.

The gain function to decide which dose to administer to new cohorts reverts back to the patient gain now. Despite the good performance of the restricted variance gain function (for permissible doses only), the ethical consideration that the dose believed to be closest to the true TD should be administered to patients is most important here. The results from the using the patient gain were not inferior to those with the restricted variance gain so there is no detrimental effect on estimation when using this gain function. In fact the patient gain is now used between cycles too, to decide whether doses administered in one cycle to a cohort should be changed for the next cycle given the current estimates of the parameters. The use of the patient gain function with the pseudo-data again forces the trial to administer dose 60 to the first cohort, and cohort 1 is fixed to remain on dose 60 for safety purposes. Therefore, intra-patient adjustments between cycles are first allowed when cohort 2 enter cycle 2.

The whole procedure is repeated 1000 times in order to obtain a mean estimate of the TD as determined by the end of the trial, and a mean estimate of the length of the trial. The precision of these estimates are investigated through looking at the reference range, the 2.5^{th} and 97.5^{th} percentiles, of the estimated TDs. Furthermore, the proportion of trials that produced an estimate within $\pm 30\%$ of the true TD ($366 mg/m^2$)

which is $(256.2, 475.8)$, is recorded to consider how often the trial was estimating a dose within a clinically defined balance of efficacy and safety. The proportion of trials stopping for each criterion is also presented.

6.2.2 Results

The results for the ICSDP incorporating intra-patient adjustments are displayed in Table 6-13 when the data is generated by the PO model or ICS model, along with the corresponding results from Tables 5-5 and 5-6 without intra-patient adjustments.

Design with adjustments			**ICSDP PO sim.**			**ICSDP ICS sim.**		
Variable			$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate			408.2	14.78		404.5	14.99	
(2.5, 97.5) percentiles of estimates			(254.0, 628.6)			(244.3, 629.1)		
97.5/2.5			**2.47**			**2.57**		
Min			190.3	7		2.7	7	
Max			805.6	20		906.3	20	
% in (TD±30%)			74.6			74.2		
Precision	Safety	Max. No.	72.3	18.3	9.4	72.9	17.1	10.0

Table 6- 13: Results from 1000 trials simulated by PO model or ICS model, escalated with patient gain, allowing intra-patient adjustments and not. TD=366mg/m^2.

Design without adjustment			ICSDP PO sim.			ICSDP ICS sim.		
Variable			$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate			381.5	14.67		371.2	15.02	
(2.5, 97.5) percentiles of estimates			(247.7, 575.3)			(243.2, 561.0)		
97.5/2.5			**2.32**			**2.31**		
Min			178.4	8		186.9	7	
Max			742.2	20		690.4	20	
% in (TD±30%)			83.0			85.5		
Precision	Safety	Max. No.	93.5	0.0	6.5	90.4	0.0	9.6

Table 6-13 cont.: Results from 1000 trials simulated by PO model or ICS model, escalated with patient gain, allowing intra-patient adjustments and not. TD=366mg/m^2.

Allowing intra-patient adjustments results in worse estimates of the TD which are overestimated. The precision of the estimates are also worse now. The proportion of trials that produce an estimate within a 30% limit of the true TD is nearly 10% less than before for both data generation models. The average trial length is quite comparable however, but fewer trials stopped due to precision compared to before with over 17% of trials for both data generation models now stopping for safety reasons as opposed to none that stopped in the original procedures.

6.2.3 Investigation of Results

The decrease in precision of the estimated target doses is reasonable in this case since fewer doses are being experimented with simultaneously. Since all patients in the trial are on the same dose, fewer observations on different doses occur, allowing less information to be obtained on the overall dose-response relationship.

The poor precision of the mean estimated TD needs some investigation. One would think that assessing and adjusting patients' doses more frequently would allow more chance to get to the true TD. Figure 6-3 shows the distribution of estimated TDs for

the ICSDP with intra-patient adjustments (blue) and without adjustments (red) when data are generated from the PO model. Figure 6-4 shows the estimated TDs when the data are generated from the ICS model. The dose values corresponding to 366±30% are displayed for reference.

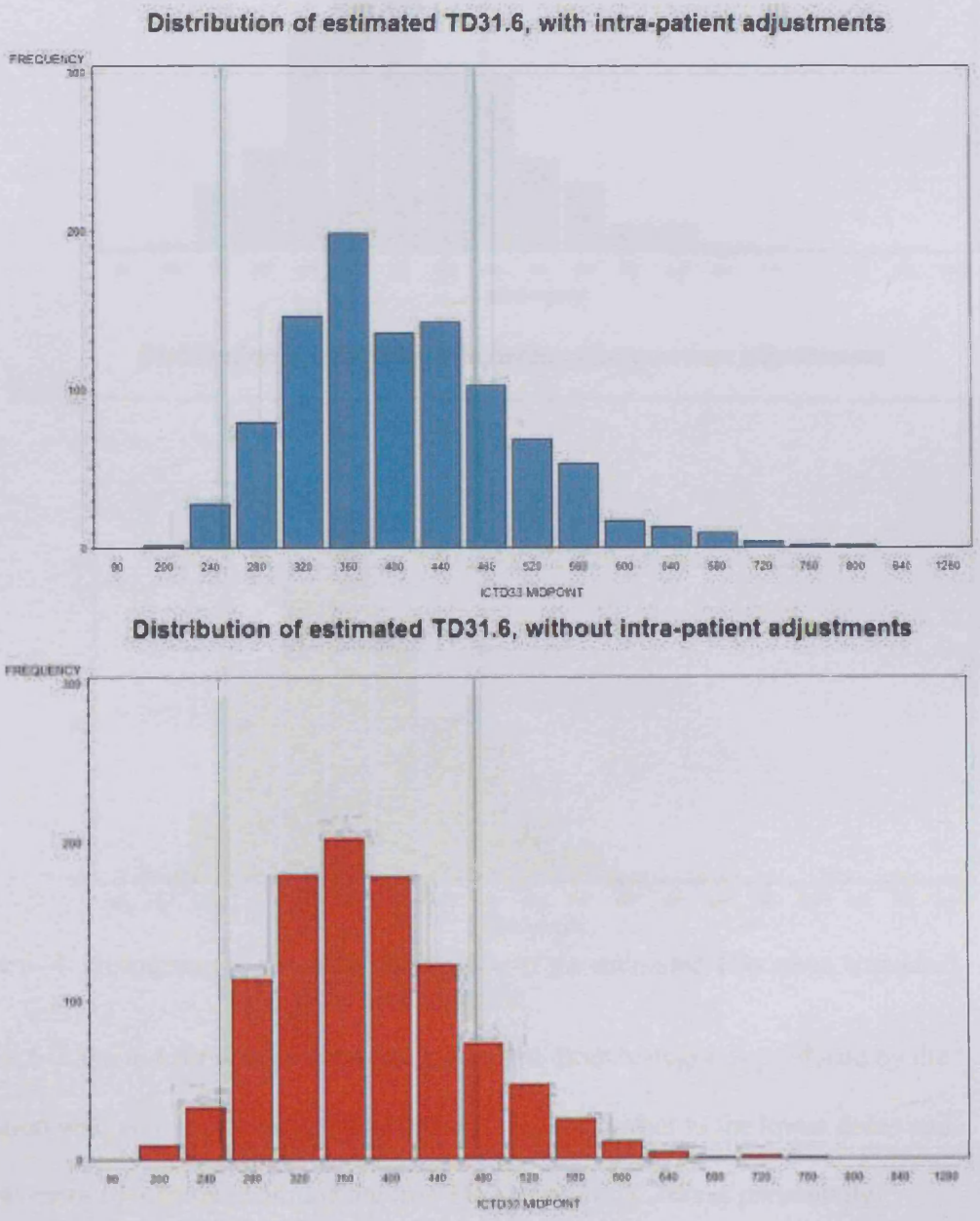


Figure 6- 3: Histograms to show the distribution of the estimated TDs when simulated by PO model.

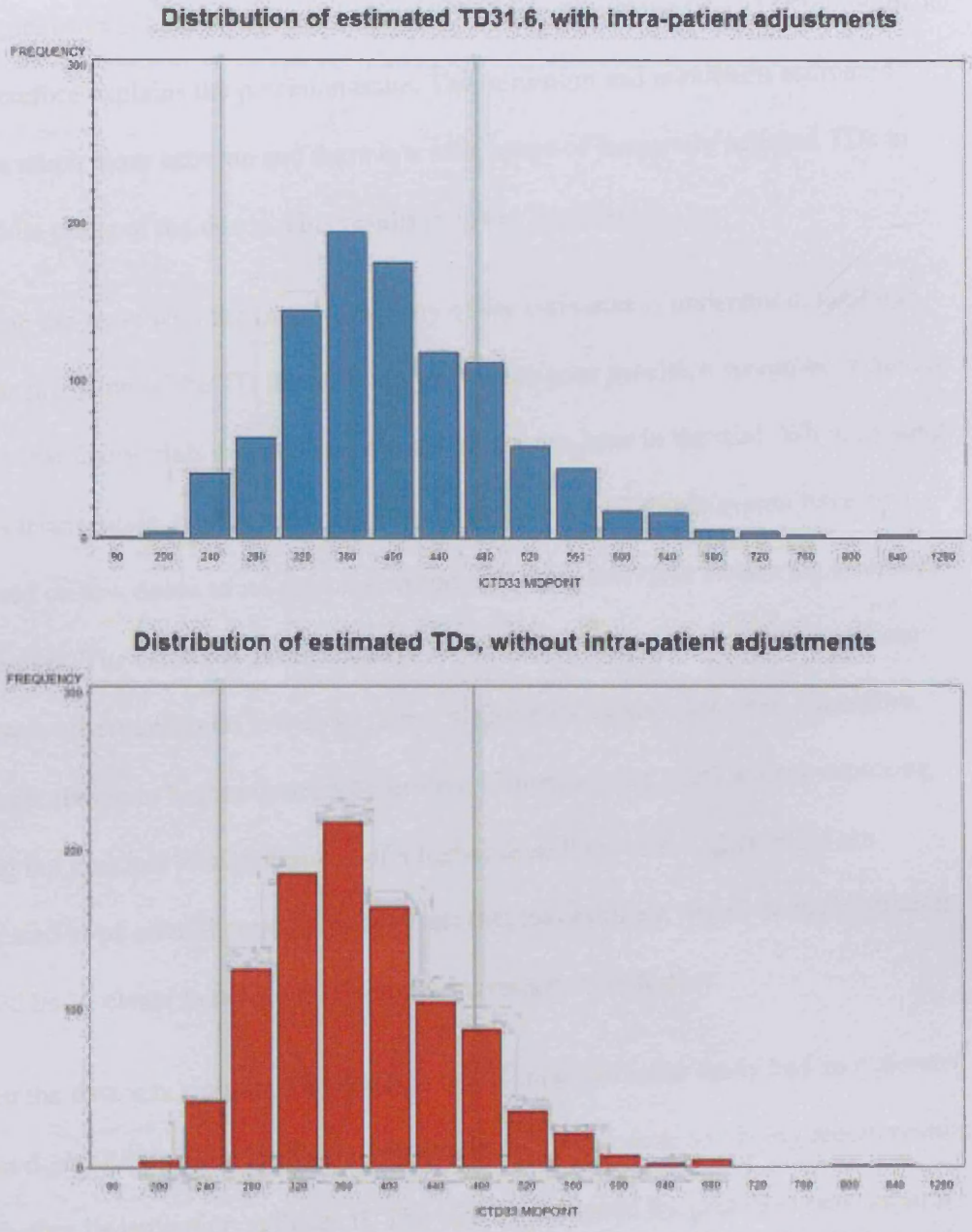


Figure 6- 4: Histograms to show the distribution of the estimated TDs when simulated by ICS model.

Figures 6-3 and 6-4 show some interesting features. Both histograms produced by the escalation with no intra-patient adjustments are skewed further to the lower doses and are also more concentrated around the true TD (366mg/m^2). This is particularly evident when looking at the trials where the data was simulated by the ICS model. The histograms produced from the intra-patient escalations show that the variability of the estimated doses is quite high, with lower and higher doses being selected as the TD

much more often. There is also not as much of a peak around the true TD (366mg/m^2). This therefore explains the precision issue. The minimum and maximum estimated TDs are much more extreme and there is a wide range of frequently selected TDs in the middle range of the doses. This results in lower precision.

Although the issue with the large variability of the estimates is understood, the issue with the precision of the TD is not. It seems that the poor precision would be related to the fact that many trials are stopping for safety reasons later in the trial. When looking at the variance gain, trials usually stop for safety because multiple events have been observed on low doses so no dose is deemed safe. This therefore causes the estimates to be lower. The estimates produced here however are higher. This is because there are fewer observations on low doses, since all patients escalate together. Therefore, faster escalation to higher doses is permitted. If there are not many events occurring during the multiple administrations of a higher dose then even higher doses are suggested to be administered. This suggests that this estimate would be higher than it should be. A closer look into one trial can investigate this further.

When the data was simulated by the ICS model, one particular study had an estimated target dose of 640.944 which would have resulted in the dose 630 being recommended for further investigation at Phase II. The trial also stopped for precision reasons so it was believed that this estimate was an accurate estimate of the target dose. The required number of cohorts to achieve this estimate was just 11. Table 6-14 shows the actual doses that were administered together with the number of DLTs observed.

Cohort and Cycle	Dose	No. Patients	No. DLTs
Cohort 1, Cycle 1,2,3	60	3 (9 observations)	0
Cohort 2, Cycle 1	200	3	0
Cohort 3, Cycle 1	300	3	0
Cohort 2, Cycle 2	300	3	0
Cohort 4, Cycle 1	420	3	1
Cohort 3, Cycle 2	420	3	0
Cohort 2, Cycle 3	420	3	0
Cohort 5, Cycle 1	630	3	2
Cohort 4, Cycle 2	630	2	0
Cohort 3, Cycle 3	630	3	0
Cohort 6, Cycle 1	420	3	0
Cohort 5, Cycle 2	420	1	0
Cohort 4, Cycle 3	420	2	0
Cohort 7, Cycle 1	420	3	0
Cohort 6, Cycle 2	420	3	0
Cohort 5, Cycle 3	420	1	0
Cohort 8, Cycle 1	630	3	1
Cohort 7, Cycle 2	630	3	0
Cohort 6, Cycle 3	630	3	0
Cohort 9, Cycle 1	630	3	0
Cohort 8, Cycle 2	630	2	0
Cohort 7, Cycle 3	630	3	0
Cohort 10, Cycle 1	630	3	0
Cohort 9, Cycle 2	630	3	0
Cohort 8, Cycle 3	630	2	0
Cohort 11, Cycle 1	945	3	2
Cohort 10, Cycle 2	945	3	1
Cohort 9, Cycle 3	945	3	0

Table 6- 14: One trial, data simulated by ICS model, escalated with intra-patient adjustments.

As can be seen, the dose escalates very quickly, to dose 630 by cohort 5. This is 2 dose levels higher than the true TD of 366 since the nearest dose levels are 300 or 420, and 630 is then above the higher of the 2 closest levels. There are very few observations on low doses: in total 18 cycles of therapy were spent on doses below the true TD, out of the total 82 patient cycles of therapy observed. Clearly this does not give much chance to observe any possible events that may occur on low doses, so immediately the estimates are likely to be higher than the true TD. 33 patient cycles were observed for dose 630 and 22 patient cycles for dose 420. This large number of observations for these particular doses contributed to the fast reduction in the

asymptotic variance and convergence to 630. Since 6 out of the 9 possible doses were administered to at least one cohort, some information was obtained throughout most of the dose range and therefore contributed to the estimation of the overall dose-response relationship. In fact once the dose 945 had been observed, the ratio of the asymptotic credible interval fell from 7.835 to 3.188 which was then deemed accurate enough to stop the trial. In order to investigate this reduction in variance further, a comparison can be made to the escalation procedure that doesn't allow intra-patient adjustments.

For two trials (one escalated with intra-patient adjustments and one not), the asymptotic variance of the log(TD) estimate was calculated. Table 6-15 shows the results for cohorts 3 and 4. Cohort 3 is the first cohort to be entered where any intra-patient adjustments could be observed, which would occur for the second cycle of cohort 2, and cohort 4 then shows further deviations. Up to cohort 3 all the dose allocations and observations were exactly the same for both procedures since cohort 1 is administered the lowest dose without dose adjustments and cohort 2 is subsequently allocated the same dose in both procedures, assuming that no DLTs are simulated on the lowest dose. The two simulated trials had similar escalations so that they could be compared directly.

Last Cohort Observed	**No Intra-Patient Escalation**		**Intra-Patient Escalation**	
	$\widehat{TD}$	*var*	$\widehat{TD}$	*var*
3 New dose=300 No DLTs	496.455	0.7273	490.817	0.6975
4 New dose=420 1 DLT for new cohort in cycle 1 (same for both)	486.142	0.57417	520.939	0.47842

Table 6- 15: Comparing Intra-Patient Escalation to No Intra-Patient Escalation.

Before cohort 3, the estimated target dose was the same for both escalations due to cohort 2 being allocated dose 200, and cohort 1 remaining on dose 60 for safety reasons. Therefore the dose to be allocated to cohort 3 was 300 for both escalations. In the Intra-Patient adjustment escalation, cohort 2 had their dose changed to 300 also, but in the original escalation procedure, cohort 2 remained on dose 200. No observations of DLTs were observed in either procedure which resulted in the estimated target doses shown above with corresponding variances. The estimated TD is higher for the original escalation (without adjustments) procedure but with larger variance. Both procedures however did result in the dose 420 being allocated to cohort 4. In the original escalation procedure, only cohort 4 are administered this dose, cohort 2 remain on dose 200 for their third cycle and cohort 3 remain on dose 300 for their second cycle. In the Intra-Patient procedure, cohorts 2, 3 and 4 are all administered dose 420 for their third, second and first cycles respectively. A DLT was observed on dose 420 for cohort 4 during cycle 1 which was common for both procedures. No other DLTs were observed. This resulted in the estimated TD and variance shown in Table 6-15. Here, the Intra-Patient escalation now has a much higher TD but with much lower variance. Since there are still multiple administrations on dose 420 in the Intra-Patient escalation, albeit during later cycles, where no DLT occurred, the DLT that did occur is not as influential. In the original escalation, 1/3 of the administrations of dose 420 resulted in a DLT hence a larger influence on the estimated TD. These estimates then result in different administrations for cohort 5. For the original escalation, dose 420 would be administered again, but for the Intra-Patient escalation, dose 630 would be administered. Although the estimated TD for the Intra-Patient escalation is nearer to dose 420, due to the dose-P(DLT) curve becoming steeper around the middle doses (doses that correspond to P(DLT)=0.5 which are

630,945mg/m^2), the new estimated probability of toxicity of dose 630 is closer to 0.316 than that for dose 420. The dose allocation is based on the probabilities associated with dose rather than dose itself, which is why 630 is allocated rather than 420. The TD estimates from the large simulation study were on average higher for the Intra-Patient escalation than the original and this may suggest why. The observations affect the estimation of the TD differently for the ICSDP with adjustments compared to the ICSDP without adjustments, due to the large differences seen in the dose administrations early on in the trial. Any DLTs that occur for higher doses early in the ICSDP with intra-patient adjustments do not have as much of a negative effect on the estimate as they should since there are also many observations of non-DLTs on the same higher doses. The variance reduces much more quickly for the Intra-Patient escalation since multiple administrations are seen simultaneously for a number of the dose levels. With this increased reduction in variance, the precision rule is achieved much sooner whilst the estimated TD is still higher than it actually should be.

6.2.4 Imposing restrictions on the intra-patient adjustments

This section looks at possible improvements to the procedure incorporating intra-patient adjustments. One example could be to investigate whether a compromise can be found to avoid sub-therapeutic dosing by allowing early cohorts some intra-patient adjustments, but stopping these between cycle adjustments once the higher doses begin to be repeated.

This is investigated by allowing a fixed number of cohorts (e.g. 3) to experience intra-patient escalation, then stopping the intra-patient adjustments once the next cohort begins (e.g. cohort 4). In actual fact this only allows three occurrences of a dose-adjustment (for cohort 2 entering cycle 2, cohort 3 entering cycle 2 and cohort 2 entering cycle 3). However, looking at table 6-3 suggests it may not be appropriate to

allow adjustments after cohort 4 begins, since this is when the estimates of the TD begin to converge more rapidly since the variance reduction begins to speed up.

The simulations are set up in exactly the same way as before (for with intra- and without intra-patient adjustments). Cohort 1 always remains on the lowest dose ($60mg/m^2$) for safety reasons. Patients in cohort 2 and 3 are then allowed to have their doses adjusted between cycles. Once cohort 4 begins, intra-patient escalation is no longer allowed and patients remain on their existing doses for the remainder of their participation in the trial. Only 100 simulations are conducted for this scenario since the use of this idea is more investigative than realistic so very precise results are not essential. However, any major deviations between procedures should still be apparent.

Based on Table 6-15, if no DLTs have been observed by the time the procedure reaches cohort 4 the dose to be administered would be $420mg/m^2$. This is the closest dose level to the true TD ($366mg/m^2$) but slightly higher, implying that lower dose levels to this would be subtherapeutic and the idea is to minimise the time that patients spend on these doses.

These results are shown in Table 6-16.

Design			ICSDP PO sim.			ICSDP ICS sim.		
Variable			$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate			403.1	15.29		389.2	15.12	
(2.5, 97.5) percentiles of estimates			(263.5, 614.6)			(265.9, 568.7)		
97.5/2.5			**2.46**			**2.14**		
Min			233.7	7		236.4	10	
Max			803.2	20		624.5	20	
% in (TD±30%)			74.0			83.0		
Precision	Safety	Max. No.	87.0	2.0	11.0	85.0	5.0	10.0

Table 6- 16: Results from 100 trials simulated by PO model and ICS model, escalated with patient gain, allowing intra-patient escalation until cohort 4 begins.

The results shown in Table 6-16 are very similar to the original investigation with the ICSDP from Chapter 5 as shown again in Table 6-13, where no adjustments are allowed, since only three adjustments are allowed here. Even with just 3 adjustments, the trial lengths are slightly increased and the TD estimates are marginally higher. So there appears to be no gain by allowing even very few adjustments and it seems that the estimates increase due to the reduced number of observations on the lower doses.

Therefore, it seems that, although it may seem unethical to keep patients on subtherapeutic doses early in the trial, the observations obtained from these low doses are essential in ensuring the model does not overestimate the TD.

Since the trial lengths were slightly shorter when allowing intra-patient adjustments throughout the trial, the method before could be reversed so that the first 3 cohorts (those on low/subtherapeutic doses) are not allowed to escalate between cycles, but once the dose has escalated to a higher dose safely, then intra-patient adjustments can be incorporated.

The results for this investigation are shown below in Table 6-17.

Design			ICSDP PO sim.			ICSDP ICS sim.		
Variable			$TD_{31.6}$	No. of Cohorts		$TD_{31.6}$	No. of Cohorts	
Mean estimate			425.1	15.25		410.6	14.81	
(2.5, 97.5) percentiles of estimates			(251.9, 701.1)			(231.6, 622.1)		
97.5/2.5			**2.78**			**2.69**		
Min			241.9	8		211.4	9	
Max			862.3	20		758.1	20	
% in (TD±30%)			63.0			68.0		
Precision	Safety	Max. No.	70.0	19.0	11.0	73.0	18.0	9.0

Table 6- 17: Results from 100 trials simulated by PO model and ICS model, escalated with patient gain, allowing intra-patient escalation after cohort 5 begins.

The results here are worse than when the intra-patient adjustments were allowed for the entire trial. This suggests that allowing intra-patient adjustments after the first few cohorts is even more detrimental to the estimation of the TD.

6.2.5 Conclusions

Incorporating intra-patient adjustments is an attractive feature and has been shown to aid convergence of the target dose, but when there is no new information or different information between cycles, the convergence is solely reliant on the dose initially administered since no additional information (such as time-changing covariates) is accruing and the dose is not permitted to change. An increased number of observations on single doses causes fast convergence of the estimated TD and a fast reduction of the variance of the estimate. This is often premature and the TD is overestimated very quickly. The suggestion of allowing just the first few cohorts to escalate between cycles to avoid sub-therapeutic dosing is a better approach than allowing adjustments throughout the trial or for later cohorts, but these results are still not as good as when no adjustments are incorporated.

It may be beneficial to use intra-patient escalation in further investigations, when there is new and accruing information available for each cycle. For example, if a marker is measured at the end of each cycle, it may provide information as to the level of response (probability of toxicity) that each patient is currently subject to. If the marker level changes, for whatever reason (due to the treatment or disease), the dose to be allocated to the next cycle may need to be adjusted accordingly.

Markers can be included as time-changing covariates, along with other baseline covariates (e.g. age, gender etc.), and a function of these covariates along with the pre-specified probability of toxicity would correspond to a patient-specific TD that could be generalised for the population, or a sub-group of the population and taken forward for recommendation at Phase II. These ideas are considered in Chapters 7 and 8.

7. Including Baseline Covariates into the Escalation Procedure

7.1. Introduction

After investigating the feasibility of intra-patient adjustments, the use of patient characteristics in the model is an important point to consider. As discussed, including more information about patients should improve the precision of the model and TD estimates as well as producing personalised TD estimates. By using patient characteristics, a more personalised dose-escalation procedure can be created which will be more ethical since the tolerability of the drug may differ across subgroups of patients, and this will need to be considered within the dose allocation methods. P(DLT) would now correspond to a randomly chosen patient from a subgroup of the population. Examples of patient characteristics include the presence and/or value of a known biomarker, or simpler ones such as age and gender. Age and gender are focused upon in this chapter.

As shown in Chapter 3, age and gender seem to have an effect on the chance of DLT. In particular, females have a higher chance of DLT than males, and younger patients are also more at risk than older patients. This will therefore form the basis for a new simulation study in which DLTs are simulated for patients dependent on their age and gender as well as dose. The ICSDP with the patient gain will again be used for the escalation procedure. The occurrence of DLTs can no longer simply be simulated by a model which just incorporates dose and cycle, so more personalised simulations need to be considered where different TDs for different patients will be produced.

7.2. Data Generation Methods

7.2.1 Covariates

Chapter 3 showed that a higher proportion of females than males experienced DLTs, particularly in cycles 1 and 2. The probability of a DLT then decreased with each cycle reasonably constantly for both males and females. The pattern adopted for generating the data is therefore based on females having a higher starting chance of toxicity, but with the probability of DLT halving for successive cycles for both males and females, as in the simulation studies in Chapters 5 and 6.

In the Postel-Vinay dataset [1] as explored in Chapter 3, patients had been categorised into four age groups, <50, 51-58, 59-65 and >65 years. There were reasonably equal numbers of patients in each category, particularly comparing the first 2 age groups to the second 2 age groups. Furthermore, the mean age of patients experiencing toxicities within each cycle varied between 50 and 63, so these four age groups seem appropriate. There was not a particularly clear pattern between the occurrence of DLT and age, however there did appear to be some relationship. The youngest group (in the first 3 cycles) nearly always experienced the highest proportion of DLTs and this proportion decreased with cycles. Apart from the first cycle, the oldest group generally had the lowest proportion of patients experiencing DLTs and this again decreased with each cycle but by a smaller proportion. The middle two age groups had a probability of DLT generally between the youngest and oldest but this probability was reasonably constant across cycles. For simulation purposes, the difference in the reduction between cycles has been ignored since it is very small and is not consistent across age groups. In order to include this, an interaction between age and cycle would need to be incorporated which would involve the use of time-dependent covariates which is outside the scope of the simulation study in this chapter. The inclusion of

time-dependent covariates will be considered in Chapter 8. The purpose of this simulation study is to show how baseline covariates are used in the dose-escalation procedure to produce a personalised estimate of the TD for subgroups of patients. Furthermore, the age categories have been simplified into two categories, <55 and ≥55. This simplification has been incorporated since the EDA from chapter 3 suggests that there were reasonably similar proportions of patients <55 and ≥55, and furthermore, the trend in P(DLT) was notably different between the lowest and highest age group, whereas the groups in the middle of the range had little difference. Therefore, the split at age 55 should reflect the difference in P(DLT) between younger and older patients. In order to incorporate more categories, the model would again become increasingly complex. The younger patients have been given a higher chance of experiencing a DLT, with the proportions experiencing a halving P(DLT) for successive cycles for all patients.

Since there was not as much of an obvious pattern in occurrence of DLTs across age groups as between males and females, the change in the probability of DLT between those aged <55 and those aged ≥55 was chosen to be smaller than that between males and females.

Based on the general pattern of DLTs across age and gender groups, a system of simultaneous equations for specific DLT probabilities and doses are created to obtain values for each of the parameters used for data generation (cycle parameters for ICS model, intercept for logistic regression model, log(dose) coefficient, age and gender parameters for both models). This is done for the "average" patient at dose 366mg/m^2 for cycles 1, 2 and 3. The "average" patient can be considered to be a patient who has the average value of the coded levels of the factor. In this case, there are 2 factors with 2 levels (age, <55 and ≥55, and gender, male and female). When the levels are coded

as 0, 1 (0=male, 1=female, 0 = <55, 1= ≥55), then if 50% of the patients are male and 50% are <55 the average patient will have value an expected value of 0.5 for each factor. Coding as 0 and 1 is the standard way to present factors with 2 levels. However in coding in this way, the baseline patient will be the patient with a value of 0 for both factors, which corresponds to a young male. It may be more appropriate to have the baseline patient as some kind of "average" patient rather than one specific type of patient. In order that the "average" patient has covariate values of 0 so to correspond to the baseline patient, the factor levels can be coded as -0.5, 0.5 (-0.5=males, <55, 0.5=females, ≥55).

Changing the coding values will change the intercept terms of the models. An investigation will be conducted to determine whether this difference actually causes an effect in the calculation of asymptotic credible intervals.

The dose of 366mg/m^2 is the target dose corresponding to a probability of a first DLT for the "average" patient in the first, second and third cycles of 0.2, 0.08 and 0.036 respectively (0.316 over 3 cycles). This corresponds to the conditional probabilities for each cycle of 0.2, 0.1 and 0.05, as in previous investigations. A dose of 799mg/m^2 corresponds to a probability of DLT in the first cycle of 0.5 for the "average" patient. The probabilities of a DLT at 366mg/m^2 in cycle 1 for the four subgroups of patients are calculated by applying a 20% reduction (approximately) in probability of DLT between females and males and a 10% reduction (approximately) between those aged <55 and those ≥55 (i.e. if the P(DLT)=0.2 for females, then P(DLT)=0.2-(0.2*0.2)=0.16 for males, and if P(DLT)=0.2 for <55s, then P(DLT)=0.2-(0.1*0.2)=0.18 for ≥55s). The actual probabilities used in the simulation are shown in Table 7-1, ordered by the least at risk to the most at risk. The TDs and the closest discrete dose levels are also displayed.

P(DLT in cycle 1 on dose 366) $= \pi_{1,366,a,g} = p_{366,a,g}(c_1)$	TD31.6	$d_{(j)} \approx TD31.6$
Old Male (a=1 or 0.5, g=0 or -0.5) = 0.17	408.12	420
Young Male (a=0 or -0.5, g=0 or -0.5) = 0.19	380.08	420
Old Female (a=1 or 0.5, g=1 or 0.5) = 0.21	352.44	300
Young Female (a=0 or -0.5, g=1 or 0.5) = 0.23	328.24	300

Table 7- 1: Probability of DLT (rounded to 2dp) and TD31.6 for each category in cycle 1.

As can be seen, the TDs for all four subgroups still lie in the interval of (300,420) as the original TD of 366 did. However, each TD is now closer to one of the two discrete dose levels of 300 or 420.

7.2.2 PO Model

The proportional odds model is of the same form as described in Chapter 5, but now incorporates the covariates into the logit link function as follows:

$$\log\left(\frac{p_{(j),a,g}(c_l)}{1-p_{(j),a,g}(c_l)}\right) = \alpha_l + \xi a + \nu g + \beta \log\left(d_{(j)}\right), \qquad (7.1)$$

where $a = 0,1$ or $-0.5, 0.5$ for the age category, and $g = 0,1$ or $-0.5, 0.5$ for the gender category.

The system of equations that require solving to obtain parameter values for the PO model (as in equation (7.1)) are shown below. When coded -0.5, 0.5, the following equations are used:

$$\log\left(\frac{p_{366,0,0}(c_1)}{1-p_{366,0,0}(c_1)}\right)=\log\left(\frac{0.2}{1-0.2}\right)=\alpha_1+\beta\log(366),$$

$$\log\left(\frac{p_{366,0,0}(c_2)}{1-p_{366,0,0}(c_2)}\right)=\log\left(\frac{0.28}{1-0.28}\right)=\alpha_2+\beta\log(366),$$

$$\log\left(\frac{p_{366,0,0}(c_3)}{1-p_{366,0,0}(c_3)}\right)=\log\left(\frac{0.316}{1-0.316}\right)=\alpha_3+\beta\log(366),$$

$$\log\left(\frac{p_{799,0,0}(c_1)}{1-p_{799,0,0}(c_1)}\right)=\log\left(\frac{0.5}{1-0.5}\right)=\alpha_1+\beta\log(799),$$

$$\log\left(\frac{p_{366,-0.5,0.5}(c_1)}{1-p_{366,-0.5,0.5}(c_1)}\right)=\log\left(\frac{0.23}{1-0.23}\right)=\alpha_1-0.5\xi+0.5\nu+\beta\log(366),$$

$$\log\left(\frac{p_{366,0.5,-0.5}(c_1)}{1-p_{366,0.5,-0.5}(c_1)}\right)=\log\left(\frac{0.17}{1-0.17}\right)=\alpha_1+0.5\xi-0.5\nu+\beta\log(366).$$

Here, ξcorresponds to the parameter associated with age, and ν the parameter associated with gender. The parameter values are shown in table 7-2.

α_1	α_2	α_3	ξ	ν	β
-11.8673	-11.4254	-11.2532	-0.1166	0.2394	1.7756

Table 7- 2: Parameter values when covariates are coded -0.5, 0.5.

These parameter values can then be used to calculate the cumulative probability of toxicity for each cycle for the different subgroups for different doses also. These probabilities are displayed in Table 7-3.

Subgroup	$p_{366,a,g}(c_1)$	$p_{366,a,g}(c_2)$	$p_{366,a,g}(c_3)$	$p_{799,a,g}(c_1)$
Old Male	0.17	0.25	0.28	0.45
Young Male	0.19	0.27	0.30	0.48
Old Female	0.21	0.29	0.33	0.52
Young Female	0.23	0.32	0.36	0.56

Table 7- 3: Cumulative probabilities of DLT for each subgroup for each cycle.

When the covariates are coded 0, 1 the baseline patient is now the young male rather than the "average" patient. Therefore, the probabilities calculated for the young male at dose 366 in cycles 2 and 3, and at dose 799 in cycle 1, from the previous system of equations are used within equation (7.1) with covariate values 0 or 1.

$$\log\left(\frac{p_{366,0,0}(c_1)}{1-p_{366,0,0}(c_1)}\right)=\log\left(\frac{0.19}{1-0.19}\right)=\alpha_1+\beta\log(366),$$

$$\log\left(\frac{p_{366,0,0}(c_2)}{1-p_{366,0,0}(c_2)}\right)=\log\left(\frac{0.27}{1-0.27}\right)=\alpha_2+\beta\log(366),$$

$$\log\left(\frac{p_{366,0,0}(c_3)}{1-p_{366,0,0}(c_3)}\right)=\log\left(\frac{0.30}{1-0.30}\right)=\alpha_3+\beta\log(366),$$

$$\log\left(\frac{p_{799,0,0}(c)}{1-p_{799,0,0}(c_1)}\right)=\log\left(\frac{0.48}{1-0.48}\right)=\alpha_1+\beta\log(799),$$

$$\log\left(\frac{p_{366,0,1}(c_1)}{1-p_{366,0,1}(c_1)}\right)=\log\left(\frac{0.23}{1-0.23}\right)=\alpha_1+\nu+\beta\log(366),$$

$$\log\left(\frac{p_{366,1,0}(c_1)}{1-p_{366,1,0}(c_1)}\right)=\log\left(\frac{0.17}{1-0.17}\right)=\alpha_1+\xi+\beta\log(366).$$

The parameter values obtained by solving these equations are given in Table 7-4 and these correspond to the same probabilities of toxicity for each subgroup as shown in Table 7-3.

α_1	α_2	α_3	ξ	ν	β
-11.9286	-11.4868	-11.3146	-0.1166	0.2394	1.7756

Table 7- 4: Parameter values when covariates are coded 0, 1.

The age and gender of each patient are simulated by two Bernoulli random variables with $p=0.5$ to represent the notion of equal proportions in each subgroup. Equation (7.1) is used with the parameter values from Table 7-4 to obtain the cumulative probabilities of DLT for each subgroup for each dose level and cycle, required for simulating DLTs:

$$p_{(j),a,g}(c_1)=\frac{\exp(\alpha_1+\xi a+\nu g+\beta\log(d_{(j)}))}{1+\exp(\alpha_1+\xi a+\nu g+\beta\log(d_{(j)}))},$$

$$p_{(j),a,g}(c_2)=\frac{\exp(\alpha_2+\xi a+\nu g+\beta\log(d_{(j)}))}{1+\exp(\alpha_2+\xi a+\nu g+\beta\log(d_{(j)}))},$$

$$p_{(j),a,g}(c_3)=\frac{\exp(\alpha_3+\xi a+\nu g+\beta\log(d_{(j)}))}{1+\exp(\alpha_3+\xi a+\nu g+\beta\log(d_{(j)}))}.$$

The method used in Chapter 5.2.2 to generate a DLT (from a Bernoulli distribution) and its corresponding cycle (from a Uniform[0,1] distribution) is adopted here again.

7.2.3 ICS Model

For the ICS model the additional covariates are incorporated into the complementary log-log link function as follows:

$$\log(-\log(1-\pi_{(j),l,a,g})) = \gamma_l + \xi a + \nu g + \theta \log(d_{(j)}), \tag{7.2}$$

where $\pi_{(j),l,a,g}$ is the conditional probability of toxicity for each cycle l on dose $d_{(j)}$ with covariate values a and g. The system of equations to be solved when the covariate values are coded -0.5, 0.5 are:

$$\log(-\log(1-\pi_{366,1,0,0})) = \log(-\log(1-0.2)) = \gamma_1 + \theta \log 366,$$
$$\log(-\log(1-\pi_{366,2,0,0})) = \log(-\log(1-0.1)) = \gamma_2 + \theta \log 366,$$
$$\log(-\log(1-\pi_{366,3,0,0})) = \log(-\log(1-0.05)) = \gamma_3 + \theta \log 366,$$
$$\log(-\log(1-\pi_{799,1,0,0})) = \log(-\log(1-0.5)) = \gamma_1 + \theta \log 799,$$
$$\log(-\log(1-\pi_{366,1,-0.5,0.5})) = \log(-\log(1-0.23)) = \gamma_1 - 0.5\xi + 0.5\nu + \theta \log 366,$$
$$\log(-\log(1-\pi_{366,1,0.5,-0.5})) = \log(-\log(1-0.17)) = \gamma_1 + 0.5\xi - 0.5\nu + \theta \log 366.$$

This leads to the parameter values displayed in Table 7-5.

γ_1	γ_2	γ_3	ξ	ν	θ
-10.0694	-10.8198	-11.5396	-0.1033	0.2129	1.4518

Table 7- 5: Parameter values when covariate values are coded -0.5, 0.5.

Table 7-6 shows the conditional probabilities and overall probability of DLT for each subgroup calculated from substituting the parameter values in Table 7-5 into equation (7.2).

Subgroup	$\pi_{j,1,a,g}$	$\pi_{j,2,a,g}$	$\pi_{j,3,a,g}$	$p_{j,a,g}(c_3)$
Old Male	0.17	0.09	0.04	0.28
Young Male	0.19	0.09	0.05	0.30
Old Female	0.21	0.11	0.05	0.33
Young Female	0.23	0.12	0.06	0.36

Table 7- 6: Conditional probabilities of DLT for each subgroup for each cycle.

When the covariate values are coded 0 or 1, the baseline patient is a young male patient. The system of equations to be solved are:

$$\log(-\log(1-\pi_{366,1,0,0})) = \log(-\log(1-0.19)) = \gamma_1 + \theta \log 366,$$
$$\log(-\log(1-\pi_{366,2,0,0})) = \log(-\log(1-0.09)) = \gamma_2 + \theta \log 366,$$
$$\log(-\log(1-\pi_{366,3,0,0})) = \log(-\log(1-0.05)) = \gamma_3 + \theta \log 366,$$
$$\log(-\log(1-\pi_{799,1,0,0})) = \log(-\log(1-0.48)) = \gamma_1 + \theta \log 799,$$
$$\log(-\log(1-\pi_{366,1,0,1})) = \log(-\log(1-0.23)) = \gamma_1 + \nu + \theta \log 366,$$
$$\log(-\log(1-\pi_{366,1,1,0})) = \log(-\log(1-0.17)) = \gamma_1 + \xi + \theta \log 366.$$

This leads to the parameter values shown in Table 7-7.

γ_1	γ_2	γ_3	ξ	ν	θ
-10.1242	-10.8746	-11.5944	-0.1033	0.2129	1.4518

Table 7- 7: Parameter values when covariate values are coded 0, 1.

The DLTs are simulated progressively for each cycle for each patient given their covariate values and dose levels as in Chapter 5.2.3.

7.3 Pseudo-data

In order to initiate the procedure, pseudo-data needs to be used. Since there are different categories of patients, pseudo-observations need to be incorporated for all categories. The pseudo-data is based on the assumption that there is no difference in the probability of DLT between different categories of patients, so that all incoming patients will be treated the same until observations are analysed in conjunction with the covariate values. The pseudo-data is set up similarly to the situation without covariates. The same set of independent Beta distributions are assigned to the probability of DLT in cycle 1 for the lowest and highest dose levels $(j) = 1, k$, as shown in Table 5-4 for each subgroup. Beta distributions are also assigned to the probability of DLT for subsequent cycles, but the parameters for these distributions are dependent on the number of toxicities observed in earlier cycles. Table 7-8 shows the calculated probabilities of DLTs for each cycle for each category of covariates for

several different doses from the ICS model used to generate data for the simulation. The doses displayed are 366, 799 and the discrete dose levels surrounding these. For reference the dose levels 60 and 1700 (lowest and highest dose levels) are displayed also to show just how pessimistic the prior information is by assigning the lowest and highest dose levels to the P(DLT) associated with the true TD and the true 50% toxic dose.

		Dose d_j					
		60	**300**	**366**	**420**	**799**	**1700**
$\pi_{j,1,a,g}$	Old Male	0.0137	0.1330	0.1735	0.2076	0.4467	0.8298
	Young Male	0.0152	0.1464	0.1904	0.2274	0.4812	0.8597
	Old Female	0.0169	0.1619	0.2100	0.2501	0.5192	0.8882
	Young Female	0.0187	0.1778	0.2300	0.2732	0.5560	0.9119
$\pi_{j,2,a,g}$	Old Male	0.0065	0.0652	0.0860	0.1040	0.2438	0.5667
	Young Male	0.0072	0.0720	0.0949	0.1147	0.2664	0.6044
	Old Female	0.0080	0.0800	0.1053	0.1271	0.2923	0.6446
	Young Female	0.0089	0.0883	0.1161	0.1399	0.3184	0.6825
$\pi_{j,3,a,g}$	Old Male	0.0032	0.0323	0.0428	0.0521	0.1272	0.3344
	Young Male	0.0035	0.0357	0.0474	0.0576	0.1400	0.3633
	Old Female	0.0039	0.0398	0.0527	0.0640	0.1549	0.3957
	Young Female	0.0043	0.0440	0.0583	0.0707	0.1703	0.4279
$p_{j,a,g}(c_3)$	Old Male	0.0232	0.2157	0.2769	0.3270	0.6348	0.9509
	Young Male	0.0257	0.2348	0.3020	0.3554	0.4727	0.9647
	Old Female	0.0286	0.2596	0.3304	0.3873	0.7046	0.9760
	Young Female	0.0316	0.2834	0.3591	0.4191	0.7511	0.9840

Table 7- 8: Probabilities of DLTs associated with different doses and dose levels from ICS data simulation model.

The pseudo-data incorporated is shown in Table 7-9.

	Covariate Category	**Dose $d_{(j)}$**	$\pi_{(j)}^0$	$n_{(j)}^0$	$t_{(j)}^0=n_{(j)}^0\pi_{(j)}^0$
ICSDP ***TTL=0.316***	For each category a=0,1, g=0,1 a=-0.5,0.5, g=-0.5, 0.5	$d_{(1)}$, cycle 1	0.2	3	0.6
		$d_{(1)}$, cycle 2	0.1	2.4	0.24
		$d_{(1)}$, cycle 3	0.05	2.16	0.108
		$d_{(k)}$, cycle 1	0.5	3	1.5
		$d_{(k)}$, cycle 2	0.2791	1.5	0.41865
		$d_{(k)}$, cycle 3	0.1473	1.08135	0.1593

Table 7- 9. Pseudo-data for all categories of patients

The pseudo-data is the same as in Table 5-6, but is now repeated for each subgroup so there is four times the amount of pseudo-data implemented compared to in Section 5-3. It can clearly be seen that the pseudo-data is likely to be very influential, since there are 12 pseudo-patients starting each dose level, and 24 pseudo-patients in total. Since there will only be a maximum number of 60 patients recruited, the minimum contribution the pseudo-data will provide is $24/84 \approx 29\%$ of the overall information. This amount will increase as the observed number of patients reduces, i.e. if the precision rule stops the trial after 15 cohorts, only 45 patients will have been recruited so the pseudo-data will then account for $24/69 \approx 35\%$ of the total information. The amount of pseudo-data to use will be investigated in later parts of this chapter.

7.4 Escalation Procedure

Once the patient characteristics and observations have been generated, the simulated escalation procedure can be carried out.

In order to begin the procedure in the same cautious fashion as has been conducted previously, pessimistic prior information needs to be used to initiate the dose allocation. This prior information is incorporated once again via the use of pseudo-data as described in Table 7-9. The lowest dose is set to correspond to the target toxicity level for all categories of patients. In doing this, all patients in the first cohort will be allocated the lowest dose possible, regardless of covariate values. In reality, there may be some prior belief as to how the tolerance of the drug may differ for different categories of patients so different starting doses could be adopted for different subgroups. However, a common safety measure is to force all patients to start at the lowest dose so this is the approach adopted here. Allowing for a covariate effect in the formulation of the pseudo-data will be incorporated in a later investigation to see whether final estimates of the TD are improved. A point for

consideration with regards to the pseudo-data is how much to include. Previously, 3 patients per dose level were used. Since there are now 4 subgroups of patients, all with a potentially different dose-response relationship, pseudo-data for each subgroup needs to be included. In the first instance, the existing method will be carried forward and 3 patients will be included for each category on each dose level as shown in Table 7-9. This is likely to be too much prior information since this will correspond to 24 patients worth of data in the prior. Since the maximum number of patients is set at 60 (20 cohorts of 3 patients) the inclusion of 24 patients in the prior information is going to be very influential since it will correspond to a minimum of 24/84 patients worth of information. This suggests that a minimum of nearly 1/3 of the information will be provided by pseudo-data. In section 7-5, different amounts of pseudo-data will be investigated, as will the exclusion of the pseudo-data from the final analysis and estimation of the TD once the trial has stopped.

Once the first cohort has been allocated the lowest dose for their first cycle, the observations of DLTs (including the pseudo-data) are analysed by the ICS model with age and gender specified as covariates. The procedure can be extended to include more covariates and continuous variables as well as factors. The parameter estimates obtained are then used along with the recorded number of patients and toxicities on each dose/cycle/age category/gender to produce estimates of the TDs for each subgroup.

In the case of no covariates the asymptotic variance of the estimate of $\log(TD_{TTL})$, where $TTL = 0.316$, was used to produce a Credible Interval (CI_U:CI_L) for $\log(TD_{TTL})$ which was then exponentiated to obtain a CI for TD_{TTL}. The ratio (R) of the exponentiated limits (CI_U:CI_L) was then used as the precision criterion by which

the trial can be stopped. In the presence of covariates, the estimate of $\log(TD_{TTL})$ is different for each category of patients, and this would result in a different asymptotic variance, CI and R for each category. One would then not be able to stop the trial based on one precision criterion. Instead, a set of precision criteria would have to be developed, all of which would have to be met in order to stop the trial for precision. Also, to find the asymptotic variance using the delta method as has been done so far, inversion of a 6x6 matrix $\left(I_E(\gamma_1,\gamma_2,\gamma_3,\xi,\nu,\theta)\right)$ along with pre and post-multiplication by a 1x6 gradient vector of $\log(TD_{TTL})$ would have to be conducted. This is not a straightforward calculation which will only become increasingly complicated as more covariates are incorporated. Furthermore, this calculation will have to be computed for every subset of the population with a different combination of covariates.

In order to include covariates in the model, an amendment to the precision stopping criterion has been considered which makes the procedure more straightforward. The calculation of $\log(TD_{TTL})$ from the ICS model with covariates is given by:

$$\log(TD_{TTL}) = \frac{1}{\theta}\left[\log\left(-\frac{\log(1-TTL)}{\left\{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right\}e^{\xi a}e^{\nu g}}\right)\right].$$

One can note from this expression that the estimate of $\log(TD_{TTL})$ is dependent on the values of the covariates (a,g). However, if one were to rearrange this expression, it can be made into a function of $\log(TD_{TTL})$ that is invariant to what set of covariates are used.

$$\log(TD_{TTL}) = \frac{1}{\theta}\log(-\log(1-TTL)) - \frac{1}{\theta}\log\left(\left\{e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right\}e^{\xi a}e^{\nu g}\right),$$

$$\log(TD_{TTL}) = \frac{1}{\theta}\log(-\log(1-TTL)) - \frac{1}{\theta}\log\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right) - \frac{1}{\theta}\log\left(e^{\xi a+\nu g}\right),$$

$$\left[\log(TD_{TTL}) + \frac{\xi a+\nu g}{\theta}\right] = \frac{1}{\theta}\log(-\log(1-TTL)) - \frac{1}{\theta}\log\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right),$$

$$F\left(\log(TD_{TTL})\right) = \left[\log(TD_{TTL}) + \frac{\xi a+\nu g}{\theta}\right],$$

$$= \frac{1}{\theta}\log(-\log(1-TTL)) - \frac{1}{\theta}\log\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right).$$

(7.3)

The delta method can then be applied to this function $F\left(\log(TD_{TTL})\right)$ to obtain an asymptotic variance for the function. Since this will not change dependent on which combination of covariates are used, only one CI and R need to be used for the precision criterion. The specific estimate of $\log(TD_{TTL})$ for each set of patients can then be found by subtracting $\xi a+\nu g/\theta$ from the function depending on values of a and g.

Using the function $F\left(\log(TD_{TTL})\right)$ as in equation (7.3) also makes the computation of the asymptotic variance much simpler. Although $I_E(\gamma_1,\gamma_2,\gamma_3,\xi,\nu,\theta)$ is still a 6x6 matrix which needs inverting, the gradient vector is now of the function $F\left(\log(TD_{TTL})\right)$ where $\frac{\partial F}{\partial \xi}, \frac{\partial F}{\partial \nu} = 0$ (where F is the function $F\left(\log(TD_{TTL})\right)$ in equation (7.3)) so the asymptotic variance of $F\left(\log(TD_{TTL})\right)$ is of the same form as the asymptotic variance of $\log(TD_{TTL})$ in the case of no covariates. The expression for the resulting variance is shown in Appendix 5. This simplification is particularly attractive when considering the fact that many covariates could now be included in this model without the expression of the asymptotic variance becoming too complicated. One

concern that may arise with the use of this function F when coding either 0, 1 and -0.5, 0.5, is that the function with covariate values equal to 0 actually corresponds to a slightly different dose. When coded as 0, 1, the estimate of F corresponds to the TD for the young males (380mg/m^2), whereas when coded as -0.5, 0.5, the estimate corresponds to the TD for the 'average' patient (366mg/m^2). Since the function with covariate values equal to 0 corresponds to different doses when coded differently, the CI may be slightly different and therefore the ratios of the CI limits may be slightly different. The estimated TDs are very similar however, and therefore any difference in the values of R should be very minimal if present at all. A simple test for this would be to analyse just the pseudo-data and compare when the values are coded as 0,1 and -0.5, 0.5. Since the pseudo-data incorporated in Table 7-9 contains no prior belief of a covariate effect, the function F is not dependent on coding values since the values of ξ and ν are equal to 0. The pseudo-data can be extended however to incorporate a prior covariate effect (details will be given later in section 7.7) where the dose of 60 is set to correspond to each subgroup's true target toxicity level associated with the dose 366mg/m^2. When this is incorporated, the prior parameters associated with this pseudo-data are found to be

$\gamma_1 = -2.8880, \gamma_2 = -3.6386, \gamma_3 = -4.3574, \theta = 0.3390, \xi = -0.1305$ and $\nu = 0.2131$ when coded -0.5, 0.5, and $\gamma_1 = -2.9429, \gamma_2 = -3.6935, \gamma_3 = -4.4122, \theta = 0.3390, \xi = -0.1305$ and $\nu = 0.2131$ when coded 0, 1. These values correspond to a TD (associated with a TTL of 0.316) of 60mg/m^2 for the average patient (with covariate values set to 0) when coded -0.5, 0.5, and 71mg/m^2 for the baseline patient (with covariate values set to 0) when coded 0, 1. The respective CIs are [10.74, 335.09] when coded -0.5, 0.5 with a ratio of 31.19 and [12.63, 393.76] with a ratio of 31.17. There is a difference in ratios, however it is very minor. Analysis of just the pseudo-data should produce a

relatively large variance, so any difference here is likely to reduce once data are observed and included. Therefore any difference should reduce further, and this will be looked at when investigating the full escalation procedures.

Once the estimate of F is obtained, the individual estimates of $p_{(j),a,g}(c_3)$ given a patients covariates of a and g are obtained and used in the patient gain function to show which dose would be best to allocate to patients based on their covariate values. The next cohort is then allocated the relevant doses (depending on what covariate values they have) and the next cycle for the first cohort is kept on the same dose. No intra-patient escalation has been allowed for this dose escalation procedure, as the investigation allowing intra-patient adjustments produced negative results when there is additional information accruing, other than the occurrence or not of DLTs. Since the covariates to include here are baseline characteristics, they will not change between cycles so there is no added benefit of allowing between cycle adjustments. Intra-patient adjustments will be considered again in Chapter 8 when considering the inclusion of time-dependent covariates. However, since different patients are allocated different doses, the information on the drug-response relationship should still be very good, even if there are more covariates to consider.

The dose escalation procedure will again continue until a stopping criterion is achieved (either safety, precision or a maximum number of patients). Estimates of the patient specific TD and the length of each trial will be obtained, along with the proportion of TD estimates that are within a 30% limit of the true TD (for each subgroup). The standardised value of the patient specific TD estimate compared to the

true patient specific TD $\left({}^{TD_{TTL}}/_{TD_{TTL}} \right)$ is also presented to see how close this value is to 1.

The first investigation, concerns the difference in coding of the age and gender covariates. The results can be compared in terms of the TD estimates, the length of the trial, the average values of R at the end of the trial and the stopping reasons. This can then be used to decide how to code the covariates.

Further investigations will then be considered, as discussed previously, to determine the role of pseudo-data in the escalation procedure, i.e. how much to use, whether to incorporate a covariate effect, whether to use it in the analysis for the final TD estimate, and also how it is created.

Once a final decision has been made as to how to incorporate the covariate values and the pseudo-data, a large simulation study will be conducted to confirm the results of this personalised ICSDP.

7.5 ICSDP with baseline covariates

7.5.1 Results

The first investigation starts with 3 patients per subgroup per dose in the pseudo-data, a total of 24 patients. This seems unreasonable to do in practice since the maximum number of patients in a trial is only 60 therefore by adding this amount of prior information is likely to overwhelm the results. However, since the initial investigation is simply to compare the coding procedure, it is used here. The comparisons between coding values are based on 100 paired datasets. The datasets are generated by both the PO and ICS models with covariate values coded either as 0,1 or -0.5,0.5. Once generated, the dataset is duplicated and the covariate values are reassigned to the

coding values that were not used for generation, i.e. the data generated by the PO model with covariate values coded as 0, 1, is duplicated and the repeated dataset has the covariate values recoded to -0.5, 0.5. The comparisons to be considered are then within generation procedure (between the paired datasets e.g. PO generated with 0,1, analysed by 0,1 or -0.5, 0.5), within data generation model (e.g. PO generated with 0,1 to PO generated with -0.5,0.5) and across data generation models but within coding generation (e.g. PO generate with 0,1 to ICS generated with 0,1).

Simulated from IC with covariate values 0, 1.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	369.66	369.42	341.55	344.16
$\widehat{TD}_{31.6}/TD_{31.6}$	0.973	0.905	1.041	0.976
Mean R	3.8818255			
Mean number of Cohorts	16.09			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	99%	-	1%	
	Covariate values coded -0.5, 0.5.			
Subgroup True TD	Young Male 380.08	Old Male 408.11	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	369.66	369.42	341.55	344.16
$\widehat{TD}_{31.6}/TD_{31.6}$	0.973	0.905	1.041	0.976
Mean R	3.8817698			
Mean number of Cohorts	16.09			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	99%	-	1%	

Table 7- 10: Results from 100 trials simulated by ICS model with covariate values 0, 1 or -0.5, 0.5.

Simulated from IC with covariate values -0.5, 0.5.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	352.18	365.23	299.52	314.35
$\widehat{TD}_{31.6}/TD_{31.6}$	0.927	0.895	0.913	0.892
Mean R	3.909659			
Mean number of Cohorts	16.16			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	96%	-	4%	
	Covariate values coded -0.5, 0.5.			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	352.18	365.23	299.52	314.35
$\widehat{TD}_{31.6}/TD_{31.6}$	0.927	0.895	0.913	0.892
Mean R	3.9096895			
Mean number of Cohorts	16.16			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	96%	-	4%	

Table 7- 10 cont.: Results from 100 trials simulated by ICS model with covariate values 0, 1 or -0.5, 0.5.

Simulated from PO with covariate values 0, 1.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	369.45	377.69	334.17	345.54
$\widehat{TD}_{31.6}/TD_{31.6}$	0.975	0.933	1.009	0.977
Mean R	3.8968865			
Mean number of Cohorts	15.64			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	-	2%	

Table 7- 11: Results from 100 trials simulated by PO model with covariate values 0, 1 or -0.5, 0.5.

	Covariate values coded -0.5, 0.5.			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	369.45	377.69	334.17	345.54
$\widehat{TD}_{31.6}/TD_{31.6}$	0.975	0.933	1.009	0.977
Mean R	3.8968472			
Mean number of Cohorts	15.64			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	-	2%	
Simulated from PO with covariate values -0.5, 0.5.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	358.10	371.60	334.35	341.86
$\widehat{TD}_{31.6}/TD_{31.6}$	0.945	0.918	1.010	0.967
Mean R	3.8936099			
Mean number of Cohorts	15.99			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	-	2%	
	Covariate values coded -0.5, 0.5.			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	358.10	371/60	334.35	341.86
$\widehat{TD}_{31.6}/TD_{31.6}$	0.945	0.918	1.010	0.967
Mean R	3.8935741			
Mean number of Cohorts	15.99			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	-	2%	

Table 7-11 cont.: Results from 100 trials simulated by PO model with covariate values 0, 1 or -0.5, 0.5.

The results for each pair of corresponding simulations (when the coding in the generation model matches the coding in the data analysed) are identical except for the value of R (Tables 7-10 and 7-11). The difference in R values can be put down to the fact that the credible interval is based on the function *F*. When the covariate values are coded -0.5, 0.5, *F* takes a lower value than when coded by 0, 1. The variance of this

function takes a lower value due to some of the terms in the variance depending on F with the rest being invariant to the change of covariate values. The credible interval limits are then slightly closer to the mean estimate of F and therefore the ratio of the limits is less. This difference, however, is only noticeable at 5 decimal places. It never results in a difference in the use of the precision stopping rule so the different values of R do not appear to affect the procedure.

Since there is no evidence of a difference in the results when the covariate values are coded 0, 1 or -0.5, 0.5, further investigations continue with the 0, 1 coding. Although it may be statistically intuitive to think of the baseline patient as the "average" patient, clinically it may be difficult to justify comparing subgroups of patients to hypothetical patients.

The differences between generating data either with parameters corresponding to coding values 0, 1 or -0.5, 0.5, can be put down to the random nature of simulation and the fact that there are only 100 simulations. While fixing a seed in the generation of the data would make the procedures exactly comparable, the random nature of the data generation across procedures does indicate the overall stability of each procedure and will highlight any obvious trends that occur due to the different methods.

The values of R remain fairly similar within data generation model but across covariate coding values used for generation, as do the average trial length and the reasons for stopping. Similar trends do appear in the TD estimates, namely that they are generally underestimated, apart from those for young females which have the lowest true TD. The TD for young females tends to be overestimated apart from in one simulation scenario, which could be due to random chance. The difference

between data generation models is also quite minimal, suggesting the ICSDP is still quite robust to model misspecification, as in Chapter 5.

The pessimistic pseudo-data used for the prior information is quite heavy and therefore becomes quite informative. This has caused the estimates to generally be lower than they should be. Comparisons of the true DLT probabilities for dose levels 60mg/m^2 and 366mg/m^2 (Table 7-8) shows that by setting dose 60 to correspond to the true TD in order to initiate a safe escalation, imposes very pessimistic information. This therefore explains the underestimation of the target doses since the pseudo-data is so heavily weighted, and therefore informative. The mean estimated TDs for the subgroups of patients all lie in a much smaller range than their true TDs. The pseudo-data was set in such a way that there was no covariate effect and all patients were treated the same. Since this prior was so informative, this trend has continued throughout the analyses and is still quite apparent in the final estimates.

7.5.2 Removing the pseudo-data from the final analysis

The final TD estimates could be recalculated without the pseudo-data included. In practice this may be an attractive option to consider since the pseudo-data is only set to initiate a cautious escalation. It does not depict actual belief so it may be appropriate to remove it once useful and insightful data has been obtained.

Initially, the pseudo-data was removed from the final dataset for each trial and the data was reanalysed stratified by trial. The calculation of the asymptotic variance however was not possible. The inclusion of the pseudo-data not only ensures that the escalation begins cautiously by putting a high DLT probability on the lowest dose, it also puts a slightly lower probability of DLT than may actually be true on the highest dose. The true overall probability of DLT for the highest dose for the baseline patient is ~0.97. Also, the conditional probability of DLT for the first cycle is ~0.87. Since

the escalation rarely reaches these high doses, when events occur at the slightly lower doses, the probabilities for the higher doses are exaggerated. This is particularly extreme when the pseudo-data is removed since there are no lower probabilities associated with the high doses included. This then causes the estimate of the conditional probabilities (particularly for cycle 1) to be equal to 1. On calculating the asymptotic variance, there are terms included that require taking the natural logarithm of $1-\pi_{(j),l,a,g}$. When $\pi_{(j),l,a,g}=1$, this logarithm tends to negative infinity hence the calculation of the variance is not possible.

In order to combat this, the pseudo-data associated with just the baseline patient (young male, age and gender covariate values = 0) can be left in. This ensures there is some extra information put onto the highest dose so that the estimated probability of toxicity does not tend to 1.

Some of the results from Tables 7-10 and 7-11 are shown again in Table 7-12, this time reanalysed after the trial has stopped with only the pseudo-data for the baseline patient.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	412.497	603.910	432.603	774.220
$\widehat{TD}_{31.6}/TD_{31.6}$	1.085	1.480	1.318	2.197
Mean R	3.961			
Mean number of Cohorts	16.16			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	96%	-	4%	

Table 7- 12: Some results from Tables 7-8 and 7-9 reanalysed without the pseudo-data.

Simulated from PO.	Covariate values coded 0, 1.			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	424.872	698.287	497.226	843.171
$\widehat{TD}_{31.6}/TD_{31.6}$	1.121	1.726	1.502	2.384
Mean R	4.0716			
Mean number of Cohorts	15.99			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	-	2%	

Table 7-12 cont.: Some results from Tables 7-8 and 7-9 reanalysed without the pseudo-data.

As can be seen, the TD estimates are now much worse than previously when the pseudo-data was included. When simulated by the ICS model, the estimate for the baseline patient (young male) is now on average overestimated by the same amount it was underestimated before. The TDs for the other subgroups are now largely overestimated, particularly for the category where there is strictly no pseudo-data information (old female, age and gender covariate values = 1). The same trend is seen when the data is simulated by the PO model. However, now even the TD for the baseline patient is overestimated by more than it was underestimated before. The mean R value is slightly increased, indicating a reduction in precision, and in the case where the data is simulated by the IC model, the average R value is greater than the precision cut-off point.

These results suggest that including the pseudo-data in the final estimates is important for multiple reasons. Firstly, not including any pseudo-data results in the mathematical issue of the calculation of the asymptotic variance being impossible, so some pseudo-data should be included to ensure that the high doses, that are unlikely to actually be experimented with, do not have exaggerated probabilities of DLT. Secondly, only including pseudo-data for certain subgroups of patients results in large overestimation of results and a reduction in the precision of these results. Since so few patients in

each subgroup are actually seen, not including any prior information about certain combinations of covariates again exaggerates results that could be obtained due to random coincidence.

7.6 Investigating the amount of Pseudo-data to use

Since it has been concluded that some pseudo-data should be included for all covariate combinations, the question is then how much pseudo-data should be included? Clearly including n=3 pseudo-patients for each patient subgroup (in this case 4 subgroups) is too much and is generally resulting in an underestimation of the TD estimates. The fact that the informative pseudo-data depicts no difference between the subgroups of patients also results in the final estimates spanning a much smaller range of doses than is actually true.

The next investigation then is to incorporate the same number of pseudo-patients as was used in the case of no covariates, i.e. n=3 pseudo-patients per dose, but to split these 3 patients between the subgroups. This will result in each of the 4 categories having n=3/4 patients starting cycle 1 on each dose.

7.6.1 Results from n=3/4 pseudo-observations per covariate category per dose

Changing the amount of pseudo data to correspond to 3 per dose level gives the following results.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	371.53	423.67	314.70	333.24
$\widehat{TD}_{31.6}/TD_{31.6}$	0.978	1.038	0.959	0.946
Mean R	523.146			
Mean number of Cohorts	9.93			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	93%	7%	-	
Simulated from PO.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	410.25	429.75	349.25	363.36
$\widehat{TD}_{31.6}/TD_{31.6}$	1.083	1.062	1.055	1.028
Mean R	5.081			
Mean number of Cohorts	10.73			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	2%	-	

Table 7- 13: Results from 100 trials simulated by ICS or PO models with n=3/4 per category for pseudo-data.

Table 7-13 shows the results with n =3/4 pseudo-patients per subgroup per dose. These results are somewhat different to what was found when n=3 pseudo-patients per subgroup per dose were used. The main point to note is the fact that the R values are bigger (particularly noticeable when simulated by the IC model) and the safety rule is used here. The TD estimates across the subgroups are more widely spread, but they are not more accurate. Generally the estimated target doses are larger than before, although those for females are less when data are simulated from the ICS model. The results from the escalations when the data is simulated by the ICS model have a particularly large average R value. This is due to the fact that in seven instances, the safety rule had to be used. Of these 7 trials, 4 of them did not proceed to cohort 2 since an event was randomly generated in cycle 1 for a patient in cohort 1 (dose 60) and because such little information was used a priori, this event overrode the prior

information and suggested that even the lowest dose was unsafe. In one of these instances, the value of R that was computed was equal to 2072.31, the other 3 gave similar R values, which explains why the average R value is so skewed. The other 3 times the safety rule was used occurred after cohort 3 and after cohort 4 where again low doses were still being administered and the first inclusion of a new subgroup resulted in a DLT in that subgroup. This again overrode the limited prior data and suggested that even the lowest doses were too toxic for that new subgroup. Because of the small amount of information, the precision of the estimates obtained was very poor, hence the extremely large average reference range ratio. This also explains the very low average number of cohorts since several stopped so early and skewed the results. When the data was simulated by the PO model, fewer events occurred on low doses. This is due to the random nature of simulation and therefore the safety rule was used less frequently. This explains the lower value of R and the slightly higher average number of cohorts.

In order to investigate how well the rest of the trials actually performed when the early events did not occur on low doses, the trials where the safety rule stopped them have been removed.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	392.47	412.73	335.885	348.44
$\widehat{TD}_{31.6}/TD_{31.6}$	1.033	1.011	1.0233	0.989
Mean R	3.763			
Mean number of Cohorts	10.53			

Table 7- 14: Results from 93 trials when generated by ICS, 98 when generated by PO. With n=3/4 per subgroup per dose for pseudo-data without trials that stopped for safety.

Simulated from PO.	Covariate values coded 0, 1.			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	398.52	434.44	353.39	368.57
$\widehat{TD}_{31.6}/TD_{31.6}$	1.049	1.065	1.077	1.046
Mean R	3.704			
Mean number of Cohorts	10.89			

Table 7-14 cont.: Results from 93 trials when generated by ICS, 98 when generated by PO. With n=3/4 per subgroup per dose for pseudo-data without trials that stopped for safety.

When data were simulated by the ICS model, the removal of the 7 trials which stopped for safety produced better estimates compared to previously and also compared to when n=3 pseudo-patients were used per subgroup (Table 7-14). When data were simulated by the PO model, the removal of the 2 trials which stopped for safety the estimates remain very similar to before. The mean R of the estimates are much improved though and the average number of cohorts required to meet the precision stopping criterion is largely reduced. It would seem that using less prior pseudo-data does create better results overall, but the issue remains that when events do happen earlier, the use of such little prior information causes the trial to stop for safety. This can happen if the first observation for a specific subgroup of patients happens to be an event, since the only other information for that patient is now of much less weight. The fact that the pseudo-data corresponds to less than 1 observation could be the reason for this. If there were at least the same amount of patients in each subgroup in the pseudo-data as in the first set of observations, the pseudo-data may not be overridden so quickly. Consideration is now given to include at least one pseudo-observation per subgroup per dose. In this setting that would correspond to 4 pseudo-patients per dose level which is still much less than the original idea of 3 per

subgroup per dose, resulting in 12 pseudo-patients per dose level, and similar to the idea of 3 per dose-level overall.

7.6.2 Results from n=1 pseudo-observations per covariate category per dose

Using 1 pseudo observation per patient subgroup per dose level gives the following results.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	402.93	403.53	350.02	355.74
$\widehat{TD}_{31.6}/TD_{31.6}$	1.060	0.989	1.066	1.009
Mean R	10.171			
Mean number of Cohorts	12.11			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	2%	-	
Simulated from PO with covariate values 0, 1.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	384.64	432.54	335.47	361.15
$\widehat{TD}_{31.6}/TD_{31.6}$	1.015	1.069	1.013	1.021
Mean Estimate of R	4.667			
Mean number of Cohorts	12.28			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	99%	1%	-	

Table 7- 15: Results from 100 trials simulated by ICS or PO models with n=1 per subgroup per dose for pseudo-data.

The results from using one patient per subgroup seem much better (Table 7-15). The use of the safety rule is observed much less here, suggesting that having one patient per covariate category does indeed stop the random occurrence of early DLTs overriding any cautious pseudo-data that has been included. There are still some occurrences of the safety rule being used which is to be expected if more than one

DLT occurs for one covariate category early on. This is very infrequent though. The TD estimates are generally similar to those found when using n=3/4 per subgroup for the pseudo-data but slightly better. The only result that seems worse from these results is the average number of cohorts required. This is slightly larger here, but the only reason that it is larger is because the trials that stopped for safety very early on due to early occurrences of DLTs no longer stop for safety so early. When compared to Table 7-14, using n=3/4 pseudo-patients per subgroup per dose but without the safety stops, the required number of cohorts is more similar, however it is still slightly larger here. This is due to the fact that slightly more pseudo-data implies that the model will not converge to the expected parameter estimates quite as quickly, and therefore not produce TD estimates that are good in such a quick time.

The results from using 1 pseudo-patient per subgroup per dose are much more promising. There is still the issue though that some trials are stopping for safety due to the overriding of the pseudo-data early on. One might want to use slightly more than one pseudo-patient per subgroup per dose. However, since one does not want to use too much pseudo-data due to the fact the estimates do not converge properly, it can be suggested to use 1.5 patients per subgroup per dose. This should still recommend that if one DLT occurs for a low dose early on, the pseudo-data is still more informative than the observed data so lower doses should still be administered. If more than one DLT is observed, although the data is now more informative, the pseudo-data still has some substantial weight rather than contributing just half of the information. This should therefore reduce the extreme probability of a DLT occurring for the lowest doses and reduce the chance of the safety rule stopping the trial for fear of the lowest dose being too toxic.

7.6.3 Results from n=1.5 per covariate category per dose

Using more than 1 pseudo-observation per patient subgroup gives the following results.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	386.71	390.15	350.84	363.94
$\widehat{TD}_{31.6}/TD_{31.6}$	1.017	0.956	1.069	1.033
Mean R	5.213			
Mean number of Cohorts	13.87			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	97%	1%	2%	
Simulated from PO with covariate values 0, 1.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	394.30	401.57	361.60	352.58
$\widehat{TD}_{31.6}/TD_{31.6}$	1.041	0.992	1.092	0.997
Mean Estimate of R	3.883			
Mean number of Cohorts	14.21			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	0%	2%	

Table 7- 16: Results from 100 trials simulated by ICS or PO models with n=1.5 per subgroup per dose for pseudo-data.

The results from the investigation with n=1.5 pseudo-patient per subgroup per dose shows that this does generally eliminate the problem of the safety rule causing the escalation procedure to cease very early on (Table 7-16).

The TD estimates here are generally very similar to when less pseudo-data was incorporated. However, the compromise of requiring slightly more cohorts is again apparent. Particularly when compared to Table 7-14. Again this is due to the increased

amount of pessimistic prior information included which reduces the speed of the parameter convergence to the expected parameter values.

There is still one occurrence of a safety stop but this is due to one of the most at risk subgroups experiencing a DLT in the first cycle of the first cohort. The first cohort contained patients from two different categories (1 older female, 2 younger males). The older female category is more at risk than the younger males and experienced a DLT during their first cycle whilst the younger males did not. The second cohort then were recruited and consisted of 1 older male, 1 younger female and 1 older female. By the second cohort, the only observations for the older male and younger female subgroups are from the pseudo-data. The observations from cohort 1 do increase the probability of a DLT for these categories, but the pseudo-data is still informative enough to ensure that the lowest dose still seems safe. However, for the older females, the observed DLT at the lowest dose suggests that it is not safe enough for this subgroup.

These results suggest that using n=1.5 for the pseudo-data does stop the safety rule from being used too easily early on in the trial, however on the occasion when there is an observed event on the only patient within a covariate category (e.g. older female) during the first cycle of the first cohort, should a patient in the second cohort be in the same covariate category then there is not enough information to suggest that the lowest dose is actually safe.

This is not unreasonable however, as in earlier investigations when no covariates were incorporated, trials still occasionally stopped for safety reasons early on when a DLT randomly occurred on a low dose.

The average length of the trial is also quite short. Although some compromise has had to be reached in order to ensure the trials do not stop for safety too often, the average number of cohorts is still similar to previous investigations despite the fact that there is slightly more prior information per dose level which could slow down the parameter convergence.

7.6.4 Conclusions

Using n=1.5 pseudo-patients per subgroup per dose level is suitable to implement as it produces reasonable TD estimates which are specific to each subgroup, even when no prior knowledge of a differing risk between covariate categories is known. It also provides enough prior information to ensure that early observations of DLTs do not override the pseudo-data and stop early for safety reasons unless completely necessary (i.e. when a DLT occurs in cycle 1 of cohort 1 and no other observations for that category are obtained).

The remaining question is then whether it is ethical to treat all patients, regardless of their baseline covariates, at the same initial dose level.

7.7 Investigating the Inclusion of a Prior Covariate Effect

This investigation looks at including a covariate effect in the pseudo-data, so that patients could start the escalation at different doses which may not necessarily be the lowest dose, dependent on their covariate levels.

In order to incorporate this, some thought needs to go into how the first dose should be allocated. The most at risk subgroup investigated is that of the younger females. Given the actual simulation parameter values, the probabilities of DLT at dose 366mg/m^2 and the closest available dose levels for the different patient subgroups are shown in Table 7-17.

	P(DLT)		
Covariate category	**300**	**366**	**420**
Young Female	0.2834	0.3591	0.4191
Old Female	0.2596	0.3304	0.3873
Young Male	0.2348	0.3020	0.3554
Old Male	0.2157	0.2769	0.3270

Table 7- 17: P(DLT) at various doses for each covariate category

The dose level 300 produces a probability of DLT closest to 31.6% for the female patients, whereas the dose level 420 produces the probability of DLT closest to 31.6% for the male patients. This therefore implies that different dose levels may well be required to be administered to different categories of patient.

In order to incorporate a covariate effect in the pseudo-data, the pseudo-data can be set so that the lowest dose 60mg/m^2 corresponds to the probability of DLT at dose 366 as shown in Table 7-17. However, given the parameters associated with the pseudo-data, the doses that would correspond to P(DLT)=0.316 would suggest that the females require a dose level below the lowest dose which is not feasible. Therefore, the dose to be administered to the first cohort should correspond to the highest P(DLT)=0.3591. This would ensure that the young female subgroup would be allocated the lowest dose 60 and the less at risk subgroups are then able to receive a slightly higher dose.

The system of equations associated with the relevant pseudo-data is shown in equation (7.4).

$$\begin{aligned}
&\log(-\log(1-\pi_{366,1,0,0})) = \log(-\log(1-0.19)) = \gamma_1 + \theta \log 60 \\
&\log(-\log(1-\pi_{366,2,0,0})) = \log(-\log(1-0.09)) = \gamma_2 + \theta \log 60 \\
&\log(-\log(1-\pi_{366,3,0,0})) = \log(-\log(1-0.05)) = \gamma_3 + \theta \log 60 \\
&\log(-\log(1-\pi_{799,1,0,0})) = \log(-\log(1-0.48)) = \gamma_1 + \theta \log 1700 \\
&\log(-\log(1-\pi_{366,1,0,1})) = \log(-\log(1-0.23)) = \gamma_1 + \nu + \theta \log 60 \\
&\log(-\log(1-\pi_{366,1,1,0})) = \log(-\log(1-0.17)) = \gamma_1 + \xi + \theta \log 60
\end{aligned} \tag{7.4}$$

This calculates the relevant parameter values based on the existing P(DLT)s at dose 366mg/m^2 as calculated in previous chapters. Those parameters are shown in Table 7-18.

γ_1	γ_2	γ_3	θ	ξ	ν
-2.9425	-3.6929	-4.14127	0.3389	-0.1033	0.2129

Table 7- 18: Parameter values associated with the pseudo-data

According to these parameters, the doses that have probability closest to P(DLT)=0.3591 are d=60mg/m^2 for females and d=120mg/m^2 for males irrespective of age.

Table 7-19 shows the pseudo-data that was implemented, based on the parameter values in Table 7-18, to initiate these dose escalations.

	Covariate Category	**Dose $d_{(j)}$**	$\pi_{(j)}^0$	$n_{(j)}^0$	$t_{(j)}^0=n_{(j)}^0\pi_{(j)}^0$
ICSDP ***TTL=0.3591***	a=0, g=1 a=-0.5 g=0.5	$d_{(1)}$, cycle 1	0.23	3	0.69
		$d_{(1)}$, cycle 2	0.1161	2.31	0.2682
		$d_{(1)}$, cycle 3	0.0583	2.0209	0.119
		$d_{(k)}$, cycle 1	0.556	3	1.668
		$d_{(k)}$, cycle 2	0.3184	1.332	0.4241
		$d_{(k)}$, cycle 3	0.1704	0.9079	0.1546
ICSDP ***TTL=0.3591***	a=1, g=1 a=0.5 g=0.5	$d_{(1)}$, cycle 1	0.21	3	0.63
		$d_{(1)}$, cycle 2	0.1053	2.37	0.2496
		$d_{(1)}$, cycle 3	0.0527	2.1204	0.1117
		$d_{(k)}$, cycle 1	0.5192	3	1.5576
		$d_{(k)}$, cycle 2	0.2923	1.4424	0.4216
		$d_{(k)}$, cycle 3	0.1549	1.0208	0.1581

Table 7- 19: Pseudo-data for all subgroups of patients with a prior covariate effect

ICSDP ***TTL=0.3591***	a=0, g=0 a=-0.5 g=-0.5	$d_{(l)}$, cycle 1	0.1904	3	0.5712
		$d_{(l)}$, cycle 2	0.0949	2.4288	0.2305
		$d_{(l)}$, cycle 3	0.0474	2.1983	0.1042
		$d_{(k)}$, cycle 1	0.4812	3	1.4436
		$d_{(k)}$, cycle 2	0.2664	1.5564	0.4146
		$d_{(k)}$, cycle 3	0.14	1.1418	0.1599
ICSDP ***TTL=0.3591***	a=1, g=0 a=0.5 g=-0.5	$d_{(l)}$, cycle 1	0.1735	3	0.5205
		$d_{(l)}$, cycle 2	0.0860	2.4795	0.2132
		$d_{(l)}$, cycle 3	0.0428	2.2663	0.097
		$d_{(k)}$, cycle 1	0.4467	3	1.3401
		$d_{(k)}$, cycle 2	0.2438	1.6599	0.4047
		$d_{(k)}$, cycle 3	0.1272	1.2552	0.1597

Table 7-19 cont.: Pseudo-data for all subgroups of patients with a prior covariate effect

Extra data was generated for the males for the first cohort at dose 120mg/m^2 and these were allocated to the first cohort.

The results of 100 trials with this pseudo-data are shown in the next section. While n=3 is shown for the pseudo-data here, the conclusion from the previous section suggest n=1.5 is much more appropriate so it will be adapted to correspond to n=1.5 for cycle 1.

7.7.1 Results

Incorporating a prior effect between patient subgroups in the pseudo-data gives the following results.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	391.02	401.01	332.352	342.05
$\widehat{TD}_{31.6}/TD_{31.6}$	1.029	0.983	1.013	0.971
Mean R	5.143			
Mean number of Cohorts	13.62			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	96%	3%	1%	
Simulated from PO.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	393.10	415.15	334.47	357.53
$\widehat{TD}_{31.6}/TD_{31.6}$	1.037	1.026	1.010	1.011
Mean R	4.254			
Mean number of Cohorts	13.77			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	97%	1%	2%	

Table 7- 20: Results from 100 trials simulated by ICS or PO models with covariate values 0, 1 with n=1.5 per subgroup per dose with a prior covariate effect for pseudo-data.

These results are similar to the results obtained without using the prior covariate effect in the pseudo-data (Table 7-20). However, the main issue here is the fact that there are an increased number of trials having to stop for safety again.

Since a different dose level is suitable for different subgroups of patients from the beginning but with little observed information, if more patients at a certain level of a factor (e.g. males) have been observed then the dose to allocate to the other level of the factor (e.g. females) may be increased without the relevant support. Since females tolerate a lower dose than males, this then causes trials to stop early for safety reasons.

To investigate this more, the same procedure (with a prior covariate effect) is used but this time with a larger amount of pseudo-data for each category again. This may not

be reasonable to do in practice, however to investigate the reason for the safety stops it is necessary.

7.7.2 Results with an increased amount of pseudo-data

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	370.13	406.40	305.07	335.76
$\widehat{TD}_{31.6}/TD_{31.6}$	0.974	0.996	0.929	0.9527
Mean R	3.887			
Mean number of Cohorts	15.77			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	98%	-	2%	
Simulated from PO.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	376.09	398.46	317.86	328.93
$\widehat{TD}_{31.6}/TD_{31.6}$	0.992	0.9847	0.960	0.930
Mean R	3.886			
Mean number of Cohorts	15.45			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	99%	-	1%	

Table 7- 21: Results from 100 trials simulated by ICS or PO models with covariate values 0, 1 with n=3 per subgroup per dose and a prior covariate effect for pseudo-data.

These results show that the estimated TDs are not as good as when there is less prior data included, even though there are no trials stopping for safety (Table 7-21).

Here the prior covariate effect is particularly influential. This causes the estimates to be lower than they should be, particularly for the patients who have a much lower target dose since such a heavy pessimistic prior is placed on those particular patients. As before when there was too much pseudo-data, the actual patients observed meant

that there was likely to be less observations for some categories of patients than pseudo-data.

7.7.3 Conclusions

It can be concluded that a prior covariate effect should not be implemented even if there may be some belief to suggest there will be a difference in drug tolerability for different subgroups of patients. When there is the same amount of pseudo-data as concluded before, the occurrence of stopping for safety is once again increased. Since there is particularly sceptical information on a certain type of patient, if those patients are observed to have experienced a DLT very early along with patients from another subgroup that don't, the prior effect is magnified and the trial is forced to stop if another patient of the most at risk categories is recruited.

7.8 Overall Conclusions

The conclusions from including baseline covariates are positive. It is possible to conduct personalised escalation procedures for different types of patients and conclude a target dose that is suitable for each type of patients when a difference in tolerability is apparent in different categories of patients.

In order to conduct the procedure, it should be suggested to begin with that there is no covariate effect acknowledged and all patients should enter the dose-escalation procedure with what is believed to be an equal chance of experiencing a DLT. Only if a difference in tolerability becomes recognised through the observed patients should a covariate effect be concluded. This is easily achievable due to the rearrangement of the link function used in this chapter allows the estimation of the TD with no covariate values which is then transformed to the subgroup specific TD. If it is deemed that no covariate effect is present, the TD without any covariate values can be concluded as the population TD.

The prior knowledge used to initiate the procedure should be enough to keep the escalation cautious to begin with, but not so much that the difference in target dose is not noticeable by the end of the trial. Therefore, it is proposed to use more than one pseudo-patient per covariate subgroup, specifically 1.5 pseudo-patients. When there is 2 or greater patients, it can take a long time for the observable data to obtain the same weight as the pseudo-data since there may be multiple combinations of covariate categories to recruit from. However, when there is only 1 pseudo-patient, the first observation of a patient from a certain covariate category can outweigh the sceptical prior information and the procedure may escalate too fast, recommending a dose that is too toxic for a certain patient.

The procedure producing the results in Table 7-16, the ICSDP using coding values of 0, 1 for the factor levels and having 1.5 patients per covariate category per dose in the pseudo-data, is repeated 1000 times to obtain some more precise results.

Simulated from IC.	**Covariate values coded 0, 1.**			
Subgroup True TD	Young Male 380.08	Old Male 408.12	Young Female 328.24	Old Female 352.44
Mean Estimate of TD31.6	366.70	397.11	327.63	351.94
$\widehat{TD}_{31.6}/TD_{31.6}$	0.967	0.981	0.989	0.995
Mean Estimate of R	5.567			
Mean number of Cohorts	14.03			
Use of Stopping Rules	Precision	Safety	Maximum No.	
	96.8%	1.7%	1.5%	

Table 7- 22: Results from 1000 trials simulated by ICS or PO models with n=1.5 for pseudo-data.

The results here are consistent with those in Table 7-16 with the estimates all being estimated very well in quite a short amount of time.

7.9 Remarks

The results of this chapter are very promising in showing how the ICSDP can be developed to include personalised dose-escalation procedures for different subgroups of patients. This simulation study considered four subgroups created by two factors at each of two levels however the methodology could be used for any number of subgroups.

A natural extension of this work would be to consider continuous covariates for inclusion in the development of personalised dose-escalation procedures. For example, age to the nearest year on a continuous scale could be included. To include a continuous covariate, it would be necessary to define a relationship between the continuous variable and P(DLT), e.g. a linear term in the complementary log-log link function. If age were to be included as a linear term, the pseudo-data could be created by using the model to obtain a P(DLT) at two given covariate values, e.g. age=40, 60. This was not possible with the data investigated in Chapter 3 since the covariate data, specifically age, was quite clustered around a small range (50-60) with a few exceptions. Furthermore, the differences in drugs and doses administered across the 38 trials made it inappropriate to develop assumptions regarding P(DLT), which is why the general trend of occurrence of first DLTs was utilised, but applied to a general setting of P(DLT)=20% for cycle 1.

The amount of pseudo-data may need some investigation when using continuous covariates, since it will not directly relate to specific subgroups. However, a good starting point would be to ensure that slightly more than 1 pseudo observation per covariate value per dose be included to ensure that the pseudo-data is not overridden too quickly.

The inclusion of continuous covariates will be considered in chapter 8, when the use of time-dependent covariates which can act as a marker for a patient's tolerance to the drug are explored. The approach for incorporating continuous covariates into the pseudo-data will therefore be described and discussed then.

8. Allowing For Lower Grade Toxicities in the Analysis of DLTs

The work conducted so far suggests that including information from later cycles enables the dose-escalation procedures to be shorter with generally more accurate and precise estimates of the TD. Incorporating baseline covariates to represent patient characteristics, such as gender and age, allows patients to receive personalised dose administrations. When used with the ICSDP to allow later cycles of therapy to be included, this still produces accurate and precise TD estimates in a short amount of time, suggesting that increasing flexibility to allow for differences between patients is not detrimental to the procedure.

The next stage is to consider if occurrences of toxicity which are less serious than DLTs, might be of use in the dose-escalation procedure. Although the requirement of the dose-escalation studies are to define a target dose which corresponds to a tolerable amount of dose limiting toxicities, one should still acknowledge the fact that lower grade toxicities (LGTs) occur more frequently. In fact, if LGTs (specifically grade 2) happen too frequently, clinicians may be encouraged to deescalate the dose or stop the patient from participating in the trial due to intuition that if multiple LGTs are occurring, the chance of a DLT is increased. P(DLT) now corresponds to a randomly chosen patient from the population but is adjusted based on the observations of LGTs as a marker for their tolerance.

8.1 Motivation

Investigation of the relationship between the occurrence of LGTs and the occurrence of DLTs was explored using data from a Phase I dose-escalation trial conducted at The Christie Hospital [17]. This Phase I study recruited patients with Hodgkin

Lymphoma in order to determine the MTD, and data on grades 1, 2 and 3/4 toxicities were recorded for the first 3 cycles of therapy. It has been identified by the German Hodgkin Study Group (GHSG) that an increased rate of grade 3/4 toxicities was associated with an improved outcome. Therefore dosing was recommended based on a dose corresponding to a certain level of toxicity rather than body surface area, as had been previously recommended. Levels of known biomarkers have been linked with tumour response and toxicity, so an exploratory aim of this study was to conclude some relationship between the biomarker and probability of toxicity (considering multiple levels of toxicity) to allow personalisation of future dosing, based on baseline levels of the biomarker.

The Christie dataset was used rather than the Postel-Vinay [1] dataset since a relationship between the number of LGTs and the occurrence of DLTs was to be investigated and quantified. With the Postel-Vinay dataset there was little information regarding the specifics of each study and there was likely to be a different relationship for each study. For the Christie study, more information was known and a specific relationship could be derived. However, the patterns of occurrences of DLTs over time obtained from the Postel-Vinay dataset (halving with successive cycles) are still used in the simulation study presented in this chapter, since they have been obtained from a much larger dataset.

The aim of the Christie study was to investigate the levels of a biomarker at different stages of therapy to see if there was a relationship between the occurrence of toxicity (mainly Gr3/4) and the level of the biomarker. Not all grade 3 toxicities were dose limiting according to the protocol.

To investigate the relationship between LGTs and DLTs, data from the first three cycles of therapy for each patient are used. If a grade 3 toxicity occurred that was specified in the protocol as not dose limiting, the toxicity was left as a non-DLT. If however a grade 3 toxicity occurred that was not excluded as a DLT in the protocol, yet it had not been recorded as a DLT, it was included as a DLT. If a DLT occurred in cycle 1 or 2, any further information from the relevant patient was disregarded as interest is still in the occurrence of the first DLT.

Table 8-1 shows the number of DLTs occurring in each cycle in the dataset.

Cycle	# of first DLTs/n patients
1	2/22=0.09
2	3/20=0.15
3	3/17=0.18

Table 8- 1: Number of patients with their first DLT in each cycle.

There is a slightly increasing rate of occurrence of a first DLT with cycle which is somewhat contradictory to what was found with the Postal-Vinay dataset. However this is a very small trial (n=22 patients) so it may not be completely reliable.

The dataset can then be split into two subsets, the patients who experienced a DLT in a given cycle and those that did not. Table 8-2 shows the total and average number of LGTs per patient in each subset in cycles prior to the specific cycle investigated. The observed number presented for cycle 2 are those observed during cycle 1, and presented for cycle 3 are those observed during cycle 1 and 2.

	DLT occurred		No DLT occurred	
Cycle	n	LG before	n	LG before
1	2/22	-	20/22	-
2	3/20	54, 18/patient	17/20	194, 11.4/patient
3	3/17	76, 25.3/patient	14/17	323, 23.1/patient

Table 8- 2: Numbers of LGTs for patients in cycles prior to the cycle of interest. For patients that did and did not experience a DLT in the specified cycle.

As can be seen, on average there is a higher number of LGTs occurring prior to the specified cycle for a patient who experienced a DLT when compared to a patient that survived the cycle without a DLT.

The LGTs (grade 1 and 2) have been grouped into one category since there are very few occurrences of grade 2 toxicities. When looking at grade 1 toxicities alone, there is a similar trend, where patients who experience a DLT in cycle 2 have on average 17.3 grade 1 toxicities prior to cycle 2 and those who don't experience a DLT in cycle 2 have on average 8.8. For cycle 3 the number of grade 1 toxicities prior to cycle 3 is 20 for patients who experience a DLT as opposed to 19.1 for those who don't.

It may be more reasonable to look at the number of LGTs occurring up until the occurrence of a DLT, as this includes LGTs that occur in the same cycle as a DLT but observed before it. Since the dates of all occurrences of toxicities are known, this is achievable.

	DLT occurred		No DLT occurred	
Cycle	**n**	**LGTs before (inc.same cycle)**	**n**	**LGTs before (inc.same cycle)**
1	2/22	16, 8/patient	20/22	249, 12/patient
2	3/20	77, 25.7/patient	17/20	409, 20.5/patient
3	3/17	116, 38.7/patient	14/17	543, 31.2/patient

Table 8- 3: Numbers of previous LGTs for patients that did and did not experience a DLT in a given cycle. Including LGTs that occurred in the same cycle.

Apart from the first cycle, the differences are magnified here. On average, many more LGTs occur previously for those patients experiencing a DLT than those who do not experience one. Clearly the comparison of patients experiencing DLTs to patients not experiencing them is biased due to the difference in follow up time. Therefore, patients who do not have a DLT are followed-up for LGTs over a longer period of time and so one might expect them to experience more. Evidently this is not the case.

Since the trends are similar when looking at LGTs occurring in cycles prior to the cycle of interest, and LGTs occurring right up to the time of the DLT or end of the cycle of interest, it can be concluded that an increased chance of a DLT is associated with a larger number of prior LGTs.

In order to simulate a realistic dataset with regards to occurrences of LGTs, some understanding of this relationship is required. Table 8-4 shows the number of LGTs (split also into grade 1 (G1) and grade 2 (G2) toxicities) in each cycle.

Cycle	G1	G2	LG (G1+G2)
1, n=22	215, 9.8/pat	49, 2.2/pat	265, 12.0/pat
2, n=20	198, 9.9/pat	40, 2/pat	238, 11.9/pat
3, n=17	186, 10.9/pat	74, 4.4/pat	260, 15.3/pat

Table 8- 4: Occurrence of LGTs in each cycle.

Table 8-4 suggests that there is a constant rate of occurrence of LGTs for the first 2 cycles which increases slightly for the last cycle. Since there is no obvious pattern, one could conclude for simplicity's sake that the rate of occurrence of LGTs remains constant throughout all three cycles of therapy.

8.2 Methodology

It is proposed to include the occurrence of LGTs as a way of predicting the occurrence of a DLT, specifically as a covariate. Patients should be able to deescalate between dosing cycles if more than the expected number of LGTs occurs or perhaps escalate further if they appear to be tolerating the drug better than expected.

Since the effect on the chance of DLT will change dependant on which cycle of therapy a patient is in and how many occurrences of LGTs they experience, the covariates relating to LGTs will be time-dependant. The covariates will only change between cycles though, which should not affect the proportional hazards assumption

adopted for the use of the ICS model. The argument used for the use of the piecewise Cox model to retain proportional hazards applies here also.

The piecewise Cox model assumes that covariates which differ in different periods of time can be still utilized if the proportional hazards assumption is still maintained within the different time periods. This theory can be extended to the ICS model, since the time periods are just the intervals, provided the covariates only change between intervals.

This is the case when looking at the number of LGTs in preceding cycles. The number of observed LGTs changes dependent on which cycle of therapy a patient is in but is considered to be constant throughout the cycle. This is an acceptable approach to take, since although the number of LGTs will change throughout the cycle, the dose adjustments and analysis can only occur at the end of each cycle. So it can be assumed for analysis purposes that the number observed previously, does not change throughout the cycle. Therefore the proportional hazards assumption is maintained within cycle and the ICS model assumptions can be upheld in this setting.

8.3 Simulation Methods

In order to simulate the occurrence of DLTs, the number of LGTs needs to be simulated first. As one would expect the probability of a DLT to increase with an increase in dose, it can be assumed that a higher number of LGTs will also occur with an increase in dose. It is reasonable to assume that as the mean number of LGTs increases with dose, so does the variance. Therefore, it is proposed that the number of LGTs in a cycle for a patient on dose level $d_{(j)}$ ($LG_{(j)}$) follows a log-normal distribution as shown:

$$\log(LG_j) \sim N\left(\log\left(\overline{LG_j}\right), sd^2\right).$$

From Table 8-4 it can be seen, that on average, the number of LGTs per patient per cycle occurring throughout the study is approximately 13 (12 in cycles 1 and 2, 15 in cycle 3). Since it can be assumed that the average dose throughout the trial is the TD, the average number of LGTs can be assumed to occur at the average dose, the TD. Therefore, in the simulation study, the rate at which toxicities occur according to dose are calculated based on a TD of 366mg/m^2. That is:

$$\overline{LG_j} = \lambda d_j,$$

$$13 = \lambda 366,$$

$$\lambda = 0.0355.$$

The mean numbers of LGTs for each discrete dose level are then calculated and displayed in Table 8-5.

$d_{(j)}$	60	120	200	300	420	630	945	1400	1700
$\overline{LG_{(j)}}$	2.13	4.26	7.10	10.65	14.91	22.37	33.55	49.70	60.35
$\log\left(\overline{LG_{(j)}}\right)$	0.76	1.45	1.96	2.37	2.70	3.11	3.51	3.91	4.10

Table 8- 5: Mean number of LGTs for each dose.

In order to determine an appropriate value for the standard deviation (sd) of the log-Normal distributions, the observed standard deviation of the number of LGTs from the Christie data is used. This is assumed to be the standard deviation associated with the target dose (366mg/m^2). Based on the properties of the Normal distribution, 95% of LGTs at the TD occur within 2 standard deviations of the average number of LGTs. From the Christie data, a standard deviation of 3 was found to correspond to the average number of LGTs observed. This therefore implies that 95% of LGTs occur in the interval $(13 \pm 6) = (7,19)$. Based on the log scale, when a standard deviation of 0.2

is used, the 95% interval is (2.1649, 2.9649) which corresponds to an interval on the original scale of (8.71, 19.39) which is reasonably similar to that observed from the EDA. Therefore the standard deviation to be used for all doses on the log scale is set to 0.2.

The number of LGTs that occur is then simulated from the log-Normal random variable for each cycle for each patient dependent on the dose administered. In the first simulation study it is assumed that there is a constant rate of occurrence for LGTs across cycles. This may not be appropriate in reality due to a reducing tolerance due to a prolonged exposure to the drug, but it is assumed here based on the data in Table 8-4. Once the number of LGTs has been generated for each cycle for each patient, the occurrence of a DLT can then be simulated. There are two approaches to this part of the simulation. First, interest lies in whether the inclusion of a time changing covariate aids the estimation of the probability of a DLT. Therefore, a model based on two time-dependent covariates (the number of LGTs during the first and second cycles) is used. Only the number of LGTs in cycles 1 and 2 are generated since these are the covariates that will allow adjusting doses after cycle 1 and after cycle 2 and there are only 3 cycles observed for DLTs. The model is given in equation (8.1).

$$\log(-\log(1-\pi_{l,(j),LG_{(j)1},LG_{(j)2}})) = \gamma_l + LG_{(j)1}\mu + LG_{(j)2}\sigma + \theta\log(d_{(j)}). \qquad (8.1)$$

When $l=1$, $LG_{(j)1}, LG_{(j)2} = 0$, since the number of LGTs in cycle 1 and 2 have not yet been observed. When $l=2, LG_{(j)2} = 0$, since the number of LGTs in cycle 2 have again, not yet been observed. The probability of a DLT associated with each cycle only depends on the LGTs in cycles strictly prior to the current cycle.

Second, since the number of LGTs have been generated to depend on dose also, interaction terms between the number of LGTs during the first and second cycles and

dose could be incorporated into the model to produce more appropriate values for the probability of a DLT. This is given by:

$$\begin{aligned}&\log(-\log(1-\pi_{l,(j),LG_{(j)1},LG_{(j)2}}))\\&=\gamma_l+LG_{(j)1}\mu+LG_{(j)2}\sigma+\theta\log(d_{(j)})+\kappa LG_{(j)1}\log(d_{(j)})+\tau LG_{(j)2}\log(d_{(j)}).\end{aligned}\tag{8.2}$$

Again, when $l=1$, $LG_{(j)1},LG_{(j)2}=0$, and when $l=2, LG_{(j)2}=0$.

The analysis model used for the dose adjustments between cycles in the dose-escalation procedure will be the same as model (8.1). Therefore, the data generation using model (8.2) will test the robustness of the procedure when there is inconsistency in the data generation and analysis models. There will be some investigation into using model (8.2) for the analysis also to see if the additional interaction terms aid the precision of the estimates.

In order to obtain suitable values for the parameters in the generation models, some assumptions have been made. For the generation based on model (8.1), at the average number of LGTs at the TD of 366, the same DLT probabilities will be set as in the simulations of Chapter 5and 6. This will correspond to the following equations:

$$\begin{aligned}\log(-\log(1-\pi_{1,366}))&=\log(-\log(1-0.2))=\gamma_1+\theta\log(366),\\\log(-\log(1-\pi_{1,799}))&=\log(-\log(1-0.5))=\gamma_1+\theta\log(799),\\\log(-\log(1-\pi_{2,366,LG_1=13}))&=\log(-\log(1-0.1))=\gamma_2+13\mu+\theta\log(366),\\\log(-\log(1-\pi_{3,366,LG_1=13,LG_2=13}))&=\log(-\log(1-0.05))=\gamma_3+13\mu+13\sigma+\theta\log(366).\end{aligned}$$

There are not enough equations to solve for the number of parameters however so further assumptions need to be made. When the number of LGTs observed in cycle 1 is 18 (nearly 50% more than the average number of 13), the probability of DLT will increase by 50% in cycle 2 to 0.15. When the number of LGTs observed in cycle 2 is 18, the probability of DLT in cycle 3 will increase by 50% to 0.075. This corresponds

to the complete system of equations that needs to be solved to obtain values for all 6 parameters:

$$\begin{aligned}
\log(-\log(1-\pi_{1,366})) &= \log(-\log(1-0.2)) = \gamma_1 + \theta\log(366),\\
\log(-\log(1-\pi_{1,799})) &= \log(-\log(1-0.5)) = \gamma_1 + \theta\log(799),\\
\log(-\log(1-\pi_{2,366,LG_1=13})) &= \log(-\log(1-0.1)) = \gamma_2 + 13\mu + \theta\log(366),\\
\log(-\log(1-\pi_{2,366,LG_1=18})) &= \log(-\log(1-0.15)) = \gamma_2 + 18\mu + \theta\log(366),\\
\log(-\log(1-\pi_{3,366,LG_1=13,LG_2=13})) &= \log(-\log(1-0.05)) = \gamma_3 + 13\mu + 13\sigma + \theta\log(366),\\
\log(-\log(1-\pi_{3,366,LG_1=13,LG_2=18})) &= \log(-\log(1-0.075)) = \gamma_3 + 13\mu + 18\sigma + \theta\log(366).
\end{aligned} \tag{8.3}$$

Solving these equations result in the parameter values shown in Table 8-6.

γ_1	γ_2	γ_3	θ	μ	σ
-10.0691	-11.9469	-13.7548	1.4518	0.0867	0.0837

Table 8- 6: Parameter values from model (8.1) used for simulation.

Note, that the intercept terms are much more different across cycles than in Chapter 5 because now they represent the effect when log(dose)=0, and there are no LGTs in cycle 1 and/or cycle 2.

Also, separate covariates have been included for the number of LGTs occurring in cycles 1 and 2. If a single covariate was used which was set equal to the total number of LGTs experienced so far, this value would change from one cycle to the next. This would not easily account for the fact that different doses may have been administered in different cycles and therefore a different number of LGTs would be expected in each cycle. In this setting, the values of the covariates do not change from one cycle to the next. Instead, additional covariates are included in later cycles. By including the covariates in this way, not only is the proportional hazards assumption maintained within cycle and the methodology of the piecewise Cox model still valid, but also the individual dose in each cycle and corresponding number of LGTs can be accounted for.

A similar system of equations can be used to solve for the parameter values when the interaction model as in (8.2) is used for generation. Two additional equations are needed, and those relating to the dose 799mg/m^2 in cycle 2 and 3 are included in a similar way to that in Chapter 5. The DLT probabilities are the same as in the setting where LGTs are not included, but at the dose of 799 they are associated with a higher number of LG toxicities than the average of 13. Given $\lambda = 0.0355$, the number of LGTs expected at dose 799 is 28. The following system of equations are obtained:

$$\begin{aligned}
\log(-\log(1-\pi_{1,366})) &= \log(-\log(1-0.2)) \\
&= \gamma_1 + \theta \log(366), \\
\log(-\log(1-\pi_{1,799})) &= \log(-\log(1-0.5)) \\
&= \gamma_1 + \theta \log(799), \\
\log(-\log(1-\pi_{2,366,LG_1=13})) &= \log(-\log(1-0.1)) \\
&= \gamma_2 + 13\mu + \theta \log(366) + \kappa 13 \log(366), \\
\log(-\log(1-\pi_{2,366,LG_1=18})) &= \log(-\log(1-0.15)) \\
&= \gamma_2 + 18\mu + \theta \log(366) + \kappa 18 \log(366) \\
\log(-\log(1-\pi_{2,799,LG_1=28})) &= \log(-\log(1-0.2791)) \\
&= \gamma_2 + 28\mu + \theta \log(799) + \kappa 28 \log(799), \\
\log(-\log(1-\pi_{3,366,LG_1=13,LG_2=13})) &= \log(-\log(1-0.05)) \\
&= \gamma_3 + 13\mu + 13\sigma + \theta \log(366) + \kappa 13 \log(366) + \tau 13 \log(366), \\
\log(-\log(1-\pi_{3,366,LG_1=13,LG_2=18})) &= \log(-\log(1-0.075)) \\
&= \gamma_3 + 13\mu + 18\sigma + \theta \log(366) + \kappa 13 \log(366) + \tau 18 \log(366), \\
\log(-\log(1-\pi_{3,799,LG_1=18,LG_2=28})) &= \log(-\log(1-0.1473)) \\
&= \gamma_3 + 28\mu + 28\sigma + \theta \log(799) + \kappa 28 \log(799) + \tau 28 \log(799).
\end{aligned} \tag{8.4}$$

The values for the parameters are shown in Table 8-7.

γ_1	γ_2	γ_3	θ	μ	σ	κ	τ
-10.0691	-11.9477	-13.7560	1.4518	0.4362	0.4249	-0.0592	-0.0578

Table 8- 7: Parameter values used for generation when an interaction is included as in model (8.2).

Since the aim of this investigation is to adjust each patient's dose between cycles depending on the number of LGTs in previous cycles, the simulated dataset should reflect the fact that patients may receive different doses in different cycles.

The occurrence of a DLT can be simulated for each dose for each patient for cycle 1 by using the Binomial distribution with P(DLT) = $\pi_{1(j)}$. If a DLT is simulated, no further cycles need to be simulated since the patient would not contribute any further to the analysis, although in reality they may still receive treatment at an adjusted dose. If a DLT is not simulated, the dose to be administered to cycle 2 is any of the possible dose levels. $\pi_{2,(j),LG_1}$ is calculated for each dose level in cycle 2 and the number of LGTs observed in cycle 1. Since the number of LGTs in cycle 1 is dependent on the dose in cycle 1, inclusion of this raw number of LGTs for the calculation of $\pi_{2,(j),LG_1}$ could bias the result. For example, if a patient moves from dose 60 to dose 200, the observed number of LGTs is associated with dose 60 in cycle 1. If this is used in the calculation for $\pi_{2,(j),LG_1}$ then it would suggest that a lower than expected number of LGTs occurred for dose 200 and therefore $\pi_{2,(j),LG_1}$ would be lower than it should be. In order to combat this, standardisation of the number of LGTs can be used to convert the number observed to the relevant scale for the new dose. This involves multiplying the raw number of LGTs by a constant which is the ratio of the new dose compared to the dose received in cycle 1. This approach seems appropriate due to the method of calculation for the average number of LGTs at each dose level. Suppose that d_h and d_j are two dose levels where $j, h \in 1,...,k$. Then

$$\overline{LG_j} = \lambda d_j,$$
$$\overline{LG_h} = \lambda d_h.$$
$$\overline{LG_j} \Big/ \overline{LG_h} = \frac{d_j}{d_h},$$
$$\overline{LG_j} = \frac{d_j}{d_h}\overline{LG_h},$$

When $j = h$, the standardisation factor reduces to 1 so the number of LGTs remains the same. This therefore suggests that the number of toxicities observed in an earlier cycle at a different dose can be used to directly estimate the number that would have been expected to occur at the new dose. A Binomial distribution is again used with P(DLT) $= \pi_{2,(j),LG_1}$, where LG_1 is the standardised number of LGTs that occurred in cycle 1 for the new dose for cycle 2. Again, if a DLT is simulated no further cycles need to be simulated. If a DLT is not simulated in cycle 2 each of the possible dose levels can be administered to cycle 3. $\pi_{3,(j),LG_1,LG_2}$ is calculated for each new dose level as well as the number of LGTs in cycle 1 and 2 which are both standardised to put them onto the relevant scale for the new dose in cycle 3. $\pi_{3,(j),LG_1,LG_2}$ is then used to simulate a DLT from a Binomial distribution.

The data can be recorded such that the first cycle only includes information on the current dose, the number of LGTs observed and the occurrence of a DLT. For the second cycle, the current dose, the dose administered to cycle 1, the number of LGTs observed in cycle 1 and 2, plus the standardised number of LGTs for cycle 1, and the occurrence of a DLT can be recorded. For cycle 3, the current dose, the doses administered to cycles 1 and 2, the number of LGTs observed in cycles 1 and 2 plus the standardised numbers, and the occurrence of a DLT is recorded. In doing this, when the relevant information is extracted from the dataset for contribution to the

escalation procedure, there are 9 possible doses to choose from for cycle 1. In cycle 2, there are 81 possible combinations of doses from cycle 1 and cycle 2, and in cycle 3 there are 729 possible combinations of doses from cycle 1, 2 and 3. This is excluding cohort 1 as all three cycles of cohort 1 will remain on the lowest dose for the same safety reasons as set out in previous investigations. By creating the data in this way, the DLT simulated in later cycles reflects the number of LGTs observed at different doses in different cycles. The overall probability of DLT across 3 cycles is given by $\pi_{1,(j)_1} + (1-\pi_{1,j_1})\pi_{2,(j)_2,LG1} + (1-\pi_{1,j_1})(1-\pi_{2,j_2,LG1})\pi_{3,(j)_3,LG1,LG2}$ which will depend on the different doses and number of LGTs for each cycle. Here, $(j)_l$ represents the dose level administered in cycle l.

8.4 Escalation Procedure

In order to incorporate LGTs into the model used for escalation purposes and also in the model for the final analysis once the procedure has stopped, there are some complications involved. First, the model would require a minimum of 6 parameters including three parameters corresponding to cycles, a log(dose) coefficient parameter and two covariate parameters. The covariates will consist of at least two linear terms for the number of LGTs occurring in previous cycles (cycle 1 and 2) and potentially two corresponding LG*dose interaction terms. This would be very computationally intensive in terms of computing the asymptotic variance of the estimated TD used for the calculation of the asymptotic credible interval required for the stopping criterion and would become even more complex should additional baseline covariates be incorporated or additional cycles be considered since the inversion of an $n \times n$ matrix is required. Second, since the covariate contribution for each cycle would be different, the terms would not factorise in the link function as they did when using baseline

covariates. The formula used to solve for the TD_{TTL} is based on equation (8.1), the model without the LG*dose interaction term, and is shown in equation (8.5).

$$TTL = 1-\exp\left(-\exp\left(\gamma_1+\theta\log\left(TD_{TTL}\right)\right)\right)\exp\left(-\left(\exp\left(\gamma_2+\mu LG_1\frac{TD_{TTL}}{d_{(j)1}}+\theta\log\left(TD_{TTL}\right)\right)\right)\right)$$
$$\exp\left(-\exp\left(\gamma_3+\mu LG_1\frac{TD_{TTL}}{d_{(j)1}}+\sigma LG_2\frac{TD_{TTL}}{d_{(j)2}}+\theta\log\left(TD_{TTL}\right)\right)\right)$$
$$=1-\exp\left[TD_{TTL}^{\theta}\left\{-\exp\left(\gamma_1\right)-\exp\left(\gamma_2+\mu LG_1\frac{TD_{TTL}}{d_{(j)1}}\right)-\exp\left(\gamma_3+\mu LG_1\frac{TD_{TTL}}{d_{(j)1}}+\sigma LG_2\frac{TD_{TTL}}{d_{(j)2}}\right)\right\}\right]. \quad (8.5)$$

When rearranging for the baseline covariate investigation the following expression is obtained as in equation (7.3),

$$\log\left(TD_{TTL}\right)=\frac{1}{\theta}\log(-\log(1-TTL))-\frac{1}{\theta}\log\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)-\frac{1}{\theta}\log\left(e^{\xi a+\nu g}\right).$$

Here it can be seen that the covariates can be factorised out since they are constant in each cycle. This kind of rearrangement in terms of the lower grade toxicities is shown in equation (8.6).

$$\log\left(TD_{TTL}\right)=\frac{1}{\theta}\log\left(-\log\left(1-TTL\right)\right)-\frac{1}{\theta}\log\left\{-\exp\left(\gamma_1\right)-\exp\left(\gamma_2+\mu LG_1\frac{TD_{TTL}}{d_{(j)1}}\right)\right.$$
$$\left.-\exp\left(\gamma_3+\mu LG_1\frac{TD_{TTL}}{d_{(j)1}}+\sigma LG_2\frac{TD_{TTL}}{d_{(j)2}}\right)\right\}. \quad (8.6)$$

As can be seen, the right hand side of equation (8.6) does not factorise into two parts, one dependent on cycle and one on covariates, since the covariates are dependent on cycle now and do not have the same value in every cycle.

The TD and the corresponding asymptotic variance of its estimate would therefore be different for each patient, and dependant on the covariate values on a continuous scale. This is not a feasible approach for Phase I studies, since it is required to conclude one or a few dose levels which corresponds to the TTL to take forward to Phase II for further investigation.

Furthermore, TD_{TTL} is on both sides of the equation but on different scales. $\log(TD_{TTL})$ is defined on the left, whereas TD_{TTL} on the linear scale is on the right hand side since it is used for scaling the number of LGTs observed on a potentially different dose. There is no analytical way to solve this, so an iterative type of approach would have to be utilized to find the solution. The true TDs calculated in this way for each individual subject given the relevant number of LGTs for each dose permutation over three cycles is shown in Figure 8-1.

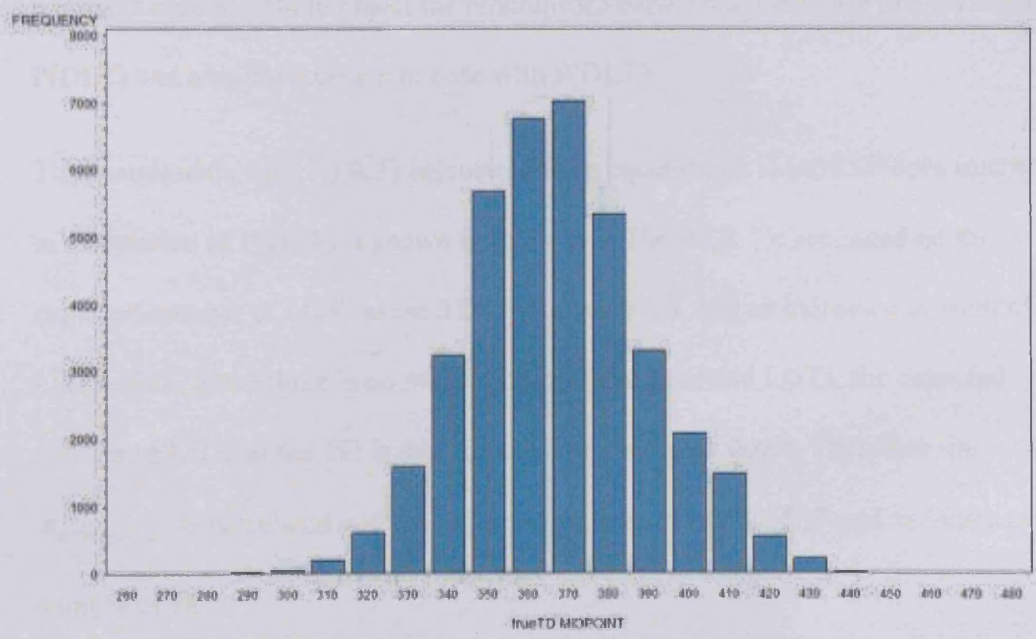


Figure 8- 1: Distribution of true Individual TDs.

The TD for a subject with LG1 = 13 and LG2 = 13 is 366mg/m^2. The average TD across all of the subjects is close to 366 mg/m^2.

A different approach is therefore considered here for the escalation procedure. This involves using the LGTs for intra-patient dose adjustments within the dose-escalation procedure in order to make personalised escalation schemes. However, only the observations of DLTs or no-DLTs are used for the analysis when it comes to administering doses to new cohorts and for the estimation of one overall TD and its asymptotic credible interval. This model is the same as in Chapter 4, equation (4.2).

As with the other procedures, the escalation will begin with the use of pseudo-data as prior information. The same total amount of pseudo-data per dose level will be used as before (n=3 as in Chapters 5 and 6) but with the inclusion of different numbers of LGTs to reflect the effect on P(DLT) when there is an increasing amount of LGTs. The systems of equations (8.3) and (8.4) are used to calculate the data generation parameter values, with and without the LG*dose interaction, and can be used to construct pseudo-data to depict the relationship between the increase in LGTs and P(DLT) and also the increase in dose with P(DLT).

The pseudo-data with P(DLT) calculated from equation (8.1) (no LG*dose interaction in calculation of P(DLT) is shown in Table 8-8. The P(DLT)s are based on the expected number of LGTs at the TD (366mg/m^2)=13, and an increased amount of LGTs of 18. Since there is no interaction between dose and LGTs, the expected number of LGTs at the TD is the expected number at all doses. Therefore the $\pi_{1(799),LG_1,LG_2}$ is calculated with an expected number of LGTs of 13 and an increased number of 18.

Obs. #	Dose, $d_{(j)}$	#LG$_1$ tox.	#LG$_2$ tox.	$\pi^0_{l(j),lG_1,lG_2}$	$n^0_{(j)l}$	$t^0_{(j)l} = \pi^0_{l(j),lG_1,lG_2} n^0_{(j)l}$
-6	$d_{(1)}$ cycle 1	0	0	0.2	1	0.2
	$d_{(1)}$ cycle 2	13	0	$\frac{1}{2}$0.2=0.1	1-0.2=0.8	0.08
	$d_{(1)}$ cycle 3	13	13	$\frac{1}{2}$0.1=0.05	0.8-0.08=0.72	0.036
-5	$d_{(1)}$ cycle 1	0	0	0.2	1	0.2
	$d_{(1)}$ cycle 2	13	0	0.1	1-0.2=0.8	0.08
	$d_{(1)}$ cycle 3	13	18	0.075	0.8-0.08=0.72	0.054
-4	$d_{(1)}$ cycle 1	0	0	0.2	1	0.6
	$d_{(1)}$ cycle 2	18	0	0.15	1-0.2=0.8	0.12
	$d_{(1)}$ cycle 3	18	18	0.1133	0.8-0.12=0.68	0.077
-3	$d_{(k)}$ cycle 1	0	0	0.5	1	0.5
	$d_{(k)}$ cycle 2	13	0	0.2791	1-0.5=0.5	0.1396
	$d_{(k)}$ cycle 3	13	13	0.1473	0.5-0.1396=0.3604	0.0531
-2	$d_{(k)}$ cycle 1	0	0	0.5	1	0.5
	$d_{(k)}$ cycle 2	13	0	0.2791	1-0.5=0.5	0.1396
	$d_{(k)}$ cycle 3	13	18	0.2151	0.5-0.1396=0.3604	0.0775
-1	$d_{(k)}$ cycle 1	0	0	0.5	1	0.5
	$d_{(k)}$ cycle 2	18	0	0.3964	1-0.5=0.5	0.1982
	$d_{(k)}$ cycle 3	18	18	0.3117	0.5-0.1982=0.3018	0.0941

Table 8- 8: Pseudo-data for the implementation of LGTs into the initiation of the ICSDP with no LG*dose interaction.

The pseudo-data with P(DLT) calculated from equation (8.2) (with LG*dose interaction in calculation of P(DLT) is shown in Table 8-9. P(DLT) at the true TD (366mg/m^2) with the expected number of LGTs (13) and an increased number of LGTs (18) are allocated to the lowest dose (60mg/m^2). P(DLT) at the 50% toxicity level (TD50, 799mg/m^2) at the expected number of LGTs (28) and an increased number of LGTs (40) are allocated to the highest dose (1700mg/m^2). 40 is chosen as the larger number of LGTs since it is near but less than the upper 97.5% percentile of the log-Normal distribution for the mean number of LGTs associated with the TD50=799, as 18 is for the TD=366. Although the expected number of LGTs at the

true TD50 (799mg/m^2) is actually 28, the value of 18 is chosen to correspond to the expected value of LGTs, and 25 is chosen for a high number of LGTs at the TD50. This still imposes an increasing amount of LGTs with dose, but does not assume that the true relationship with LGTs and dose is known.

Obs. #	Dose, $d_{(j)}$	#LG$_1$ tox.	#LG$_2$ tox.	$\pi^0_{l(j),LG_1,LG_2}$	$n^0_{(j)l}$	$t^0_{(j)l} = \pi^0_{l(j),LG_1,LG_2} n^0_{(j)l}$
-6	$d_{(1)}$ cycle 1	0	0	0.2	1	0.2
	$d_{(1)}$ cycle 2	13	0	$\frac{1}{2}$0.2=0.1	1-0.2=0.8	0.08
	$d_{(1)}$ cycle 3	13	13	$\frac{1}{2}$0.1=0.05	0.8-0.08=0.72	0.036
-5	$d_{(1)}$ cycle 1	0	0	0.2	1	0.2
	$d_{(1)}$ cycle 2	13	0	0.1	1-0.2=0.8	0.08
	$d_{(1)}$ cycle 3	13	18	0.075	0.8-0.08=0.72	0.054
-4	$d_{(1)}$ cycle 1	0	0	0.2	1	0.6
	$d_{(1)}$ cycle 2	18	0	0.15	1-0.2=0.8	0.12
	$d_{(1)}$ cycle 3	18	18	0.1133	0.8-0.12=0.68	0.077
-3	$d_{(k)}$ cycle 1	0	0	0.5	1	0.5
	$d_{(k)}$ cycle 2	18	0	0.2791	1-0.5=0.5	0.1396
	$d_{(k)}$ cycle 3	18	18	0.1473	0.5-0.1396=0.3604	0.0531
-2	$d_{(k)}$ cycle 1	0	0	0.5	1	0.5
	$d_{(k)}$ cycle 2	18	0	0.2791	1-0.5=0.5	0.1396
	$d_{(k)}$ cycle 3	18	25	0.2236	0.5-0.1396=0.3604	0.0805
-1	$d_{(k)}$ cycle 1	0	0	0.5	1	0.5
	$d_{(k)}$ cycle 2	25	0	0.4151	1-0.5=0.5	0.2076
	$d_{(k)}$ cycle 3	25	25	0.3375	0.5-0.1982=0.3018	0.0941

Table 8- 9: Pseudo-data for the implementation of LGTs into the initiation of the ICSDP with LGT*dose interaction.

8.5 Scenarios

There are a number of different elements included in this procedure. Firstly there is the generation of the LGTs. The LGTs are generated as increasing with dose independently for each cycle. The extension to this is to generate the number of LGTs in cycle 2 dependent on the number in cycle 1. This is done by allocating the mean of

the log-Normal distribution for the generation of LGTs for cycle 2, as the number of LGTs observed in cycle 1. The number observed in cycle 1 is scaled dependent on the dose in cycle 1 to the dose in cycle 2.

Second, there is the model used for generating the DLTs. There may be an interaction term for LGTs and dose in the calculation of P(DLT) or not.

Next is the use of the pseudo-data. The number of LGTs may be assumed to be increasing with dose or not (Tables 8-8 or 8-9).

Then there is the model used for the intra-patient dose adjustments, incorporating the number of LGTs in cycle 1 and 2 to make decisions on which dose a patient should receive in the next cycle. This could incorporate the interaction term or not.

Finally there is the model used for the overall analysis of DLTs, which does not depend on the number of LGTs, and is used to allocate the dose for cycle 1 of the next cohort, and also for the precision criterion of the stopping rules. This model is the model used in Chapter 4 (equation (4.2)).

Different combinations of these elements can be incorporated to investigate the robustness of the procedure when different assumptions are made.

Table 8-10 shows the different scenarios to be investigated, including which choice of each element is adopted.

Scenario	Generation of LGTs	Generation of DLTs	Pseudo-data	Intra-Patient Analysis Model
1	Cycle 1 indep. Cycle 2. LGTs incr. with dose.	Interaction of LGTs and dose in P(DLT). Equation (8.2)	Incr. LGTs with dose. Table 8-9.	No Interaction in P(DLT). Equation (8.1)
2	Cycle 1 indep. Cycle 2. LGTs incr. with dose.	Interaction of LGTs and dose in P(DLT). Equation (8.2)	Incr. LGTs with dose. Table 8-9.	Interaction in P(DLT). Equation (8.2)
3	Cycle 1 indep. Cycle 2. LGTs incr. with dose.	Increasing LGTs with dose. No Interaction of LGTs and dose in P(DLT). Equation (8.1)	Incr. LGTs with dose. Table 8-9.	No Interaction in P(DLT). Equation (8.1)
4	Cycle 1 indep. Cycle 2. LGTs incr. with dose.	Interaction of LGTs and dose in P(DLT). Equation (8.2)	LGTs not incr. with dose. Table 8-8.	No Interaction in P(DLT). Equation (8.1)
5	Cycle 2 dependent on Cycle 1. LGTs incr. with dose.	Interaction of LGTs and dose in P(DLT). Equation (8.2)	Incr. LGTs with dose. Table 8-9.	No Interaction in P(DLT). Equation (8.1)
6	Cycle 2 dependent on Cycle 1. LGTs incr. with dose.	LGTs not included in P(DLT). Equation (4.2)	Incr. LGTs with dose. Table 8-9.	No Interaction in P(DLT). Equation (8.1)
7	Cycle 2 dependent on Cycle 1. LGTs incr. with dose.	LGTs not included in P(DLT). Equation (4.2)	LGTs not incr. with dose. Table 8-8.	No Interaction in P(DLT). Equation (8.1)

Table 8- 10: Scenarios for investigation.

Scenario 1, is the basic ICSDP incorporating LGTs. This will incorporate a dose*LG interaction in the data generation to reflect the increasing number of LG toxicities with dose in the calculation of P(DLT). The analysis model used for intra-patient adjustments in the procedure will not incorporate this interaction term. The pseudo-data will also reflect the concept of an increasing number of LGTs with dose and their

interaction. Scenario 2 is the same as Scenario 1, but includes the interaction term in the analysis model for intra-patient adjustments. This is done to look at the procedure under perfect conditions, where the analysis model matches the generation model perfectly. The pseudo-data will also reflect the increasing rate of LGTs with dose and the interaction in the calculation of P(DLT). The dataset generated for these two instances will have the number of LGTs occurring in cycle 2, independent of those that have occurred in cycle 1.

The next part of the investigation will look at how the procedure performs when there are further discrepancies between the data generation and analysis.

In Scenario 3, the dataset will not incorporate LG*dose interaction in the calculation of P(DLT), nor will the analysis model for the intra-patient adjustments. The pseudo-data does increase with dose and there is also an increasing rate of occurrence of LGTs with dose. This may then suggest that P(DLT) increases with dose and the number of LGTs linearly, without taking into account the expected increase in LGTs with dose despite the fact that the procedure will begin by allowing the LGTs to increase with dose.

Some further investigation into the pseudo-data will be incorporated in Scenario 4. The rate of occurrence of LGTs does not increase with dose in the pseudo-data. However it will increase in the data generation methods, and there will be a LG*dose interaction in the calculation of P(DLT). The analysis model will also still not include an interaction term.

The next focus of these investigations will be to try to make the generated data even more realistic. In Scenario 5 the dose*LG interaction terms are included in the calculation of P(DLT). There will also be a relationship between the number of LGTs

in cycle 2 and 1. It seems reasonable that, since LGTs can be considered as an indicator for tolerance, the occurrence of LGTs in subsequent cycles would depend on the occurrence of LGTs observed before.

Based on the inclusion of some dependence between LGTs in cycles 1 and 2, some further assumptions can be tested. If a relationship between dose and LGTs is apparent, this relationship between dose and LGTs may prompt assumptions that there is a relationship between LGTs and P(DLT). If this assumption is erroneous, the procedure should be able to detect no relationship and allowing for intra-patient adjustments based on LGTs should not be beneficial, as in Chapter 6. Scenario 6 uses the data generated as in Chapter 5, where P(DLT) is calculated based on cycle and dose only. No relationship between LGTs and P(DLT) is incorporated despite a relationship between LGTs and dose. All subjects therefore have the same individual TD of 366. The procedure will still be conducted such that the model with LGTs will be used for the intra-patient adjustments despite no relationship, and a relationship between LGTs and P(DLT) will be included in the pseudo-data also. Finally, in Scenario 7, the pseudo-data can also be adapted to depict no relationship between LGTs and P(DLT) despite an increasing rate of LGT occurrence with dose. This investigates how the procedure performs with no initial suggestion of LGT and P(DLT) relationship.

For all scenarios, the model used to analyse the occurrence of DLTs alone is that used in Chapter 5, defined in equation (4.2).

8.6 Results

The results presented are initially the same as in previous chapters. The mean estimated TD (corresponding to the overall P(DLT)) along with the 95% reference range and the ratio of the reference range limits is calculated. The average number of

cohorts required to achieve the estimate of TD is presented along with the minimum and maximum number of cohorts observed. The percentage of trials stopping for each of the stopping criteria is also shown.

Furthermore, individual TDs are estimated after the trials have stopped based on the precision of the overall TD. The individual TDs are estimated through an iterative procedure where the dose (from 1 to 1700) producing $p_{j,LG_1,LG_2}(c_3)$ closest to 0.316 given the observed number of LGTs is concluded to be the patient specific TD. The mean individual TD across all patients in all simulated trials is displayed. The estimated individual TD is divided by the true individual TD, calculated from parameter values obtained from equation (8.1) or equation (8.2) depending on which was used for the data generation. The average ratio over all patients and simulations is then presented. Furthermore, the estimated individual TD can be divided by the overall estimated TD (associated with DLTs). This is again averaged over all patients and all simulations. The estimated overall TD for a given trial is divided by 366 (the true population TD) and averaged over all simulations and finally the individual estimated TD is divided by 366 and averaged over all patients and simulations.

8.6.1 ICSDP incorporating LG toxicities

The results from incorporating the occurrence of LGTs into the ICSDP when an interaction term is included in the data generation model, but not the analysis model (Scenario 1) are shown in Table 8-11.

Design			ICSDP		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			323.58		13.73
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(205.3, 482.4) **2.35**		
Min			155.37		8
Max			618.18		19
% in (TD±30%)			82.1		
Precision	Safety	Max No.	100.0	0.0	0.0

Table 8- 11: Results from Scenario 1, 1000 trials simulated by the ICS model.

When incorporating the interaction between dose and LGTs into the calculation of P(DLT) for the data generation, the estimated TD associated with DLTs is somewhat lower than the true 366. However, the variability of this estimate is very good with a very high proportion of trials producing an estimate of the TD within a 30% limit of the true TD. The expected number of cohorts required is very low and much less than the original ICSDP (Table 5-8). Also here, it is very important to note that all of the trials stop for precision suggesting that targeting doses to patients' individual needs aids the estimation of the dose-response relationship, possibly because so many different dose levels are investigated.

Table 8-12 displays a summary of the individual TDs as estimated at the end of the trial. The individual TDs are compared to the TD associated with DLTs (data analysed without LGTs) and also the true individual TDs.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_i}$	Mean estimated individual TD per trial	365.40	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	323.58	
$\overline{\left(\widehat{TD}_i / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.14	>1 - Individual TDs are higher than TD associated with DLTs
$\overline{\left(\widehat{TD}_i / TD_i\right)}$	Estimated individual TD as a proportion of the true individual TD	0.994	~1 - Estimated individual TDs are very near to true TDs
$\overline{\left(\widehat{TD}_i / 366\right)}$	Estimated individual TD as a proportion of 366	0.998	~1 - Estimated individual TDs are very near to 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.884	<1 - Estimate TD associated with DLTs are less than 366.

Table 8- 12: Mean Estimates associated with individual target doses after 1000 trials from Scenario 1.

As can be seen, the mean individual TD for a trial when averaged across the 1000 simulations is almost exactly the 366 as used for simulation, and this is confirmed when comparing the mean estimate of the individual TD to 366. Furthermore, the individual TD mean estimate compared to the true individual TD estimate is almost 1 suggesting that including LGTs in the estimation of individual TDs is extremely effective and can be used to estimate a TD for an individual person very efficiently.

The individual TD as a proportion of the TD associated with just DLTs, along with comparing the mean individual TD to the overall TD shows that the overall TD associated with just the DLTs is much less. Since there are different TDs associated with different tolerabilities the overall TD underestimates the TD in order to be acceptable for those patients with lower tolerabilities. Some explanation for the reduction in the estimates could be a result of the specific dose levels administered. Since the dose range is not spread equally, it is much more likely to move between lower doses since the P(DLT) associated with each of these doses is not too different.

If the P(DLT) changes for the current dose administered, it is likely to not change too much so a dose near to the current dose will likely be administered for later cycles. If the nearest dose level to that currently being used is some distance away, it may be that the dose with P(DLT) nearest to the TTL will still be the same dose, even if it is now believed to be too toxic or sub-therapeutic. If patients are then kept on lower doses than are perhaps suitable, it is more likely that DLTs will be observed on the lower doses. The estimates of the TD will then be skewed towards this end. To quantify this, the differences between doses (in mg/m^2) and P(DLT) (where P(DLT) is calculated by the equation (4.2)) for adjacent dose levels are shown in Table 8-13.

Dose	60	120	200	300	420	630	945	1400	1700
E(P(DLT))	0.03	0.07	0.15	0.25	0.37	0.57	0.78	0.93	0.97
mg/m^2 diff.		60	80	100	120	210	315	455	300	
P(DLT) diff		0.04	0.08	0.10	0.12	0.20	0.21	0.15	0.04	

Table 8- 13: Differences in mg/m^2 and P(DLT) between adjacent dose levels.

As can be seen, it would be a much larger jump to adjust doses upwards for doses from $420mg/m^2$. Although the differences between P(DLT) level off at the high doses, these doses are particularly toxic anyway where one would expect to observe a DLT so the likelihood of adjusting to a higher dose here is very low. One could produce LGTs to suggest their P(DLT) at dose 420 is lower than 37%, but since the next dose (630) has on average a 20% higher chance of DLT, the reduction may not be enough to suggest adjusting to a dose that much higher. Therefore a subject may remain on the dose for which they are less likely to observe a DLT. If a DLT does still occur on this lower dose, since there is still a reasonable probability of this happening, then despite the number of LGTs observed, this will lead to underestimation of the true probability of DLT.

Considering intermediate dose levels for the higher doses may combat this underestimation. However, it is not always feasible to suggest that these doses would be made/administered at multiple higher doses since they would be believed to be too toxic anyhow. It is reasonable however to recommend a slightly lower dose than the true maximum tolerated dose for further investigation. Later phases of development often use slightly lower doses than the MTD in order to demonstrate efficacy in doses that have an even lower chance of toxicity, so this is an acceptable compromise to accept for enabling targeted dosing of patients and reducing the chance of over/under-exposure.

In order to investigate whether there is an issue with regard to allowing an interaction term for LGTs and dose in the calculation of P(DLT) in the generation of the data, but escalating between cycles according to the model without interaction, the inclusion of the interaction in the escalation model is investigated (Scenario 2). Since this is just an investigative scenario, only 100 simulations will be performed.

Design			**ICSDP**		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			261.84		19.97
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(189.3, 354.4) **1.87**		
Min			77.63		18
Max			400.78		20
% in (TD±30%)			52.0		
Precision	Safety	Max No.	5.0	0.0	95.0

Table 8- 14: Results from Scenario 2, 100 trials simulated by the ICS model.

As can be seen there is a large reduction in the mean estimate of the TD with a very large increase in the average number of cohorts observed. This is due to the fact that more parameters have to be estimated. This causes a much slower escalation. With

patients escalating between cycles more slowly, the accrued dose information that is obtained across cycles is not as informative, so for the analysis of DLTs alone more of the lower doses are observed with greater variability in each cycle, causing the estimation of the TD associated with just DLTs to be worse.

The individual TDs are summarised in Table 8-15.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_i}$	Mean estimated individual TD per trial	345.14	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	261.84	
$\overline{\left(\widehat{TD}_i / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.33	>1 - Individual TDs are greater than TD associated with DLTs
$\overline{\left(\widehat{TD}_i / TD_i\right)}$	Estimated individual TD as a proportion of the true individual TD	0.918	<1 - Estimated individual TDs are less than true TDs
$\overline{\left(\widehat{TD}_i / 366\right)}$	Estimated individual TD as a proportion of 366	0.943	<1 - Estimated individual TDs are less than 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.715	≪1 - Estimated TD associated with DLTs are much less than 366.

Table 8- 15: Mean Estimates associated with individual target doses after 100 trials from Scenario 2.

The individual estimated TDs are fairly good in comparison with those from Scenario 1 (Table 8-12). The matching of the data generation model and analysis model implies that the personalized procedures are very effective. However the individual TDs are not estimated any better than in Table 8-12, potentially due to the increased number of parameters required to be estimated. This combined with the obviously lower overall TD suggests that using the interaction in the analysis model is not beneficial.

8.6.2 Testing the procedure when there are contradictions between data generation and analysis

Allowing for no interaction between dose and LGTs in the calculation of P(DLT) in the data generation, despite an increasing rate of LGTs with dose and the interaction incorporated in the pseudo-data, Scenario 3, produces the results in Table 8-16.

Design			**ICSDP**		
Variable			$TD_{31.6}$	No. of Cohorts	
Mean estimate			312.21	13.54	
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(208.9, 450.8) **2.16**		
Min			157.14	9	
Max			561.44	20	
% in (TD±30%)			81.1		
Precision	Safety	Max No.	99.7	0.0	0.3

Table 8- 16: Results from Scenario 3, 1000 trials simulated by the ICS model.

The estimate of the TD is somewhat lower than the true 366 but with good precision. The expected number of cohorts required is very low and much less than the original ICSDP. A very high proportion of trials stop for precision and a large number of estimates lie within a 30% limit of the true TD.

The TD estimate itself is lower than in Table 8-11 (Scenario 1), but this is due to the fact that the interaction between dose and LGTs is not considered in the calculation of P(DLT). LGTs are simulated at a much greater rate when the dose is higher but this is not considered in the calculation of P(DLT) from the complementary log-log link function. So when a higher number of LGTs occur at a higher dose, the effect on P(DLT) is 2-fold. It increases once with the dose and again with the number of LGTs. Therefore DLTs will be simulated with a higher probability than is reasonable. When DLTs are analysed independently at the end of the escalation, the estimated dose that corresponds to the TTL is lower than expected.

Some summary results for the individual TDs are shown in Table 8-17.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_i}$	Mean estimated individual TD per trial	356.81	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	312.21	
$\overline{\left(\widehat{TD}_i / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.154	>1 - Individual TDs are greater than TD associated with DLTs
$\overline{\left(\widehat{TD}_i / TD_i\right)}$	Estimated individual TD as a proportion of the true individual TD	0.970	~1 - Estimated individual TDs are very near to true TDs
$\overline{\left(\widehat{TD}_i / 366\right)}$	Estimated individual TD as a proportion of 366	0.975	~1 - Estimated individual TDs are very near to 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.853	<1 - Estimate TD associated with DLTs are less than 366

Table 8- 17: Mean Estimates associated with individual target doses after 1000 trials from Scenario 3.

These results show that the average individual TD is very close to the true 366 despite the overall TD associated with DLTs being so low. The ratio of the estimated individual TDs compared to the overall TD associated with DLTs confirms that the individual TDs are generally larger. The ratio of the estimated individual TDs compared to the true individual TDs show that the estimated individual TD is very close to the subject's true TD, suggesting that allowing personalised escalations is very efficient at targeting therapy to suit a specific subject even when a dose*LG interaction is not used in the calculation of P(DLT).

The next investigation involves using pseudo-data that does not necessarily match the generation of the data (Scenario 4). Although the data reflects an increasing rate of LGTs with dose and the interaction is incorporated in the calculation of P(DLT), this is not incorporated in the pseudo-data or analysis model. Suggesting the assumptions

made about the data, prior to commencing the procedure, are inconsistent with the observed data.

Design			ICSDP		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			345.26		14.14
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(218.31, 526.37) **2.411**		
Min			172.55		8
Max			719.30		19
% in (TD±30%)			83.8%		
Precision	Safety	Max No.	100%	0%	0%

Table 8- 18: Results from Scenario 4, 1000 trials simulated by the ICS model.

The results in Table 8-18 are actually better than the first investigation (Scenario 1, Table 8-11) where an increase in dose results in an increasing rate of LGTs in the pseudo-data (Table 8-9). In this scenario, there is no suggestion of an interaction between LGTs and dose in the calculation of P(DLT) in the pseudo-data. The model used for analysing the DLTs does not incorporate the occurrence of LGTs so when there is no difference across dose levels in the pseudo-data, the parameter estimates converge more quickly, despite the increasing rate of LGTs with dose in the data generation model. Therefore the estimated TD associated with DLTs is closer to the true value of 366. It is still slightly underestimated due to the increasing rate of LGTs not being reflected in the occurrence of DLTs, so P(DLT) increases linearly with dose and LGTs. A lower dose would therefore be tolerated.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_i}$	Mean estimated individual TD per trial	361.62	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	345.26	
$\overline{\left(\widehat{TD}_i / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.0703	>1 - Individual TDs are greater than TD associated with DLTs
$\overline{\left(\widehat{TD}_i / TD_i\right)}$	Estimated individual TD as a proportion of the true individual TD	0.983	~1 - Estimated individual TDs are very near to true TDs
$\overline{\left(\widehat{TD}_i / 366\right)}$	Estimated individual TD as a proportion of 366	0.988	~1 - Estimated individual TDs are very near to 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.943	<1 - Estimate TD associated with DLTs are less than 366

Table 8- 19: Mean Estimates associated with individual target doses after 1000 trials from Scenario 4.

The results associated with restricting the pseudo-data (as in Table 8-8) show that the mean individual TD is very near the true value of 366. Here it is not quite as good as in Table 8-12 since the interaction between dose and LGTs is not incorporated in the pseudo-data despite being incorporated in the data generation. Not allowing the interaction of the rate of LGT occurrence and dose in the pseudo-data slows down the convergence of the parameter estimates associated with the LGTs in the intra-patient escalation model (equation (8.1)). The individual TDs are still estimated very well though. The ratio of the estimated individual TD compared to the estimated DLT TD is closer to 1, suggesting that the two estimates are more similar. This is an attractive feature since it suggests that the overall DLT TD is estimated closer to the individual TD, so it becomes more reasonable to allow the analysis of just DLTs in order to predict a population TD. In reality, the reason for this is because the estimation of the DLT TD is better due to the matching assumptions in the pseudo-data and the analysis

model used for intra-patient adjustments. The individual TD is also estimated slightly worse due to the mis-matching, resulting in a narrower gap between the different TDs.

The overall DLT TD is estimated better when the pseudo-data doesn't impose too many relationships between LGTs and dose as the parameter estimation is less restricted. So if the focus is indeed to obtain an overall DLT TD to recommend for further investigation, it may be useful to incorporate this feature.

8.6.3 Making the data more realistic and further testing of assumptions

The results in Table 8-20 are from Scenario 5 which show how the procedure performs when an interaction between LGTs and dose is included in P(DLT) for the data generation and the pseudo-data, and the number of LGTs occurring in cycle 2 depends on the number occurring in cycle 1. This dependence of LGTs in cycle 2 and 1 depict a more realistic relationship.

Design			**ICSDP**		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			321.80		14.03
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(206.4, 478.3) **2.317**		
Min			146.56		9
Max			628.17		20
% in (TD±30%)			82.9%		
Precision	Safety	Max No.	99.1%	0%	0.91%

Table 8- 20: Results from Scenario 5, 1000 trials simulated by the ICS model.

Table 8-21 shows the corresponding individual TD results.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_i}$	Mean estimated individual TD per trial	368.94	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	321.80	
$\overline{\left(\widehat{TD}_i / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.157	>1 - Individual TDs are greater than TD associated with DLTs
$\overline{\left(\widehat{TD}_i / TD_i\right)}$	Estimated individual TD as a proportion of the true individual TD	1.010	~1 - Estimated individual TDs are very near to true TDs
$\overline{\left(\widehat{TD}_i / 366\right)}$	Estimated individual TD as a proportion of 366	1.008	~1 - Estimated individual TDs are very near to 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.879	<1 - Estimate TD associated with DLTs are less than 366

Table 8- 21: Mean Estimates associated with individual target doses after 1000 trials from Scenario 5.

The results in Table 8-21 are very similar to those obtained in Table 8-12, since the only difference is the number of LGTs in cycle 2 which will be a very minor change. The next set of results show how the procedure would work if the assumption of LGTs affecting the P(DLT) is made erroneously, based on an increasing relationship of LGTs with dose.

The data is generated with an increasing rate of LGTs with dose, but the calculation of P(DLT) is as in equation (4.2) and does not contain LGTs, so all patients have an equal individual TD of 366mg/m^2. The pseudo-data here is set to reflect the misspecified assumption of increasing P(DLT) with increasing number of LGTs (Table 8-9).

Design			ICSDP		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			340.20		13.95
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(211.8, 501.4) **2.368**		
Min			171.13		8
Max			660.45		20
% in (TD±30%)			83.06		
Precision	Safety	Max No.	97.96%	-	2.04%

Table 8- 22: Results from Scenario 6, 1000 trials simulated by the ICS model.

These results produce very reasonable estimates for the TD with good precision and a high proportion in a clinically meaningful range of the true TD, and with again quite few cohorts on average. The proportion of trials stopping for precision is nearly 100%, with very few stopping due to reaching the maximum cohort. The individual results are shown in Table 8-23.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_i}$	Mean estimated individual TD per trial	372.72	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	340.20	
$\overline{\left(\widehat{TD}_i / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.103	>1 - Individual TDs are greater than TD associated with DLTs
$\overline{\left(\widehat{TD}_i / TD_i\right)}$	Estimated individual TD as a proportion of the true individual TD	1.032	~1 - Estimated individual TDs are very near to true TDs
$\overline{\left(\widehat{TD}_i / 366\right)}$	Estimated individual TD as a proportion of 366	1.018	~1 - Estimated individual TDs are very near to 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.930	<1 - Estimate TD associated with DLTs are less than 366

Table 8- 23: Mean Estimates associated with individual target doses after 1000 trials from Scenario 6.

The average estimated individual TD is slightly larger than the true 366 but is still quite close, as shown by the average ratio of the estimated individual TD to the true individual TD.

Although there should be better estimation of the individual TD since all patients have the same TD, the reason for the estimate not being better can be put down to the fact that there is still an allowance for patients to change doses between cycles due to their observed number of LGTs, which does increase with dose. Since the data associated with LGTs outweighs the amount of information obtained on DLTs for different doses, it seems the model still relies on the occurrence of LGTs as an indicator for P(DLT). Indirectly, an increase in LGTs implies a reduced tolerance to higher doses, which then implies a higher P(DLT) for each dose level, so although the LGTs do not directly affect the calculation of P(DLT), there is some confounding relationship with dose. This may be more apparent due to the inclusion of a relationship between LGTs and P(DLT) in the pseudo-data.

Table 8-24 shows the results when the assumption is removed from the pseudo-data as in Table 8-8.

Design			**ICSDP**		
Variable			$TD_{31.6}$		No. of Cohorts
Mean estimate			341.38		13.61
(2.5, 97.5) percentiles of estimates **97.5/2.5**			(211.9,520.8) 2.458		
Min			141.20		8
Max			757.92		20
% in (TD±30%)			83.23%		
Precision	Safety	Max No.	97.46%	0%	2.54%

Table 8- 24: Results from Scenario 7, 1000 trials simulated by the ICS model.

The average TD is marginally better when removing the relationship between LGTs and P(DLT) in the pseudo-data. And it is also achieved in a slightly shorter time, due to the quicker convergence of parameters when not pre-specified to correspond to pessimistic observations, but the estimates are produced with slightly worse precision. The proportion of TDs in the 30% limit of the true TD is almost identical as to when the relationship was specified in the pseudo-data, as is the proportion of trials stopping for precision and due to reaching the maximum cohort.

	Explanation	Mean from 1000 trials	Interpretation
$\overline{\widehat{TD}_t}$	Mean estimated individual TD per trial	380.26	
$\overline{\widehat{TD}}$	Mean Estimated TD associated with DLTs	341.38	
$\overline{\left(\widehat{TD}_t / \widehat{TD}\right)}$	Estimated individual TD as a proportion of TD associated with DLTs	1.118	>1 - Individual TDs are greater than TD associated with DLTs
$\overline{\left(\widehat{TD}_t / TD_t\right)}$	Estimated individual TD as a proportion of the true individual TD	1.055	~1 - Estimated individual TDs are very near to true TDs
$\overline{\left(\widehat{TD}_t / 366\right)}$	Estimated individual TD as a proportion of 366	1.039	~1 - Estimated individual TDs are very near to 366
$\overline{\left(\widehat{TD} / 366\right)}$	Estimated TD associate with DLTs as a proportion of 366.	0.933	<1 - Estimate TD associated with DLTs are less than 366

Table 8- 25: Mean Estimates associated with individual target doses after 1000 trials from Scenario 7.

Table 8-25 shows that the results for the estimated individual TDs are actually slightly worse here which is surprising since the pseudo-data starts the procedure assuming that all patients have the same TD regardless of the number of LGTs. One can assume that since there is still a confounding effect of LGTs and P(DLT) through dose, the procedure still attempts to determine a relationship, but now there is no initial suggestion for this it does so with less targeted precision than before.

8.7 Conclusions

The investigation into incorporating LGTs into the dose escalation procedure, but analysing and making decisions based on just the occurrence of DLTs is very promising.

The basic assumption used implies that P(DLT) increases with dose and LGTs, and accounts for the fact that a different rate of LGTs will occur for different doses. The model used for analysis and intra-patient escalation purposes, is the model with the number of LGTs incorporated, but without the LGT and dose interaction, despite the data being generated with an interaction. This model is easier to implement and understand which is critical when suggesting using this model in practice. When incorporating the interaction term in the analysis model for intra-patient adjustments, the results are not improved, suggesting the more complicated model is not necessary to implement. The model used for determining an overall TD is the model without LGTs, as in Chapter 5. The results from implementing these basic assumptions show that the overall TD (associated with DLTs) can be produced quite well, although usually underestimated. The individual TDs however can be predicted extremely well. These estimates are obtained in a shorter amount of time than seen in previous investigations of the use of the ICSDP and generally with higher precision, resulting in a larger amount of trials stopping for precision and producing results within a clinically meaningful range.

When investigating scenarios that conflict with the basic assumptions, very few differences arise. When the data is generated with no interaction term, the results for the overall TD are generally worse, since a higher expected number of LGTs are observed with increasing doses, but this is not taken into account when calculating P(DLT), so there is a detrimental effect and the estimated TD is lower.

When the pseudo-data doesn't implement a relationship on the occurrence of LGTs with dose and therefore no interaction with dose in the calculation of P(DLT), despite the data being generated with such an interaction, the results are better for the overall estimated TD, but marginally worse for the individual estimated TDs. The overall TD is estimated better due to the improved escalation procedure found from the matching assumptions between pseudo-data and analysis model, but the individual TDs do not reflect the actual individual tolerance as well, although still produces very good results.

When a more extreme assumption is violated, that being that there is no relationship between LGTs and P(DLT) in the data, despite dose having a relationship with both variables, the results are still reasonably robust. The overall TD is still produced very well with very good precision and in a similar amount of time also. The individual TDs are estimated slightly worse now, due to the added complexity of estimating parameters for which there is prior suggestion to estimate, but no evidence to support it. Looking at the results from a greater perspective they are still estimated very well. The ability to conclude that LGTs do not affect P(DLT) is not achieved here, but this is down to the confounding relationship of an increasing rate of LGTs with dose, and an increasing P(DLT) with dose.

It seems that the best scenario would be to match the assumptions incorporated in the pseudo-data, to those used for the analysis models, i.e. no interaction of LGTs with P(DLT), so no increasing rate of LGTs with dose. Even when the data itself does not match these assumptions as such, the procedure is robust enough to provide good results, which are then enhanced due to the matching assumptions.

A final point to consider is the fact that in general, the average estimated individual TD is much closer to the true population TD of 366 than the TD associated with analysing DLTs alone. While the analysis of just DLTs is important to include for the stopping criteria and also to allocate a dose to new, incoming cohorts of patients, once the trial has stopped it would be appropriate to analyse all information for each patient with the model incorporating LGTs as in equation (8.1). The average individual TD can then be calculated as the dose to recommend for further investigation. This is a very efficient use of the data which incorporates more information regarding the overall tolerance of the drug.

9. Conclusions and Remarks

9.1. Overall Conclusions

Comparison of the ICSDP to an existing method (LRDP1) and a compromise between the existing method and one which incorporates more cycles of therapy (LRDP3), shows that the ICSDP performs better overall than either of the other two. The TD estimates are usually either comparable or better than those from the other procedures and this estimation is largely invariant to model misspecification. The main benefit of the ICSDP however, is that it generally requires fewer cohorts to obtain these comparable estimates, leading to a shorter trial. The trials are also more likely to stop for precision of the estimated TD than those using the other procedures and very rarely stop for safety reasons. The ability to cope with non-informative censoring is one of the original benefits of the use of the ICS model, and this is confirmed in the setting of Phase I dose-finding studies. Informative censoring, due to patients being more likely to withdraw when near to experiencing a DLT due to perhaps the occurrence of lower grade toxicities, affects all procedures similarly due to the reduced amount of information on how dose affects the chance of DLTs. The results produced by all procedures are still biased and will recommend doses that are too high. This is a different issue that would need further investigation to that conducted in this investigation.

The patient gain function is the most ethical one to use with the ICSDP. Since patients are being treated at what is believed to be the TD, the estimate of the TD will converge to the true TD when dosing is concentrated around this estimate (Shen and O'Quigley [18]). When compared to other gain functions that aim to maximise information (e.g. the variance gain function), the reduction in information is not so

great for the patient gain function. It is more important to consider the welfare of the subjects receiving the treatment, which in this case are cancer patients.

One of the attractions of using the ICSDP is the ability to change doses between cycles for a patient (intra-patient dose adjustments). It is an ethical approach since patients would not be exposed for too long to a dose that is deemed sub-therapeutic or overly-toxic. However, when no additional information is known about the patient when progressing to later cycles, the incorporation of intra-patient adjustments makes estimation of the TD much worse. More patients being dosed at the same dose level at any one time encourages the estimates of the TD to converge much quicker. However fewer doses will actually have been used over multiple time-points, so the precision of these estimates is very reduced. The idea of intra-patient dose adjustments should be considered when information is accrued with time, such as the occurrence of LGTs in earlier cycles.

Including covariates that will result in different target doses for each patient is a difficult concept to consider, especially since this is the first phase of clinical investigation so the main aim is to provide a dose, or a few doses, to recommend for later investigation. In practice, a TD can be created for each individual patient, but this is only useful when the probability of a DLT is affected by the value of the covariate. Baseline covariates that will not change and are known before a patient is administered the drug can be taken into consideration when dosing. It has been shown that incorporating such covariates into the ICSDP is possible, and the procedure can be adapted to produce multiple recommended doses.

The main issue that arises when attempting to produce multiple TDs is the construction of the pseudo-data. When just one TD is to be estimated, the equivalent

of two cohorts worth of information (1 cohort at both the lowest and the highest doses) can be incorporated, which results in less than 10% of the total information accrued if the trial were to reach its maximum size. However, when multiple TDs are being estimated, 2 cohorts worth for each subgroup will overwhelm the data and prevent proper convergence of the TD estimates. More than one pseudo-observation should be included at each dose level for each subgroup associated with a different TD, in order to ensure the first few observations do not cause the escalation to either proceed at an unsafe rate, if no events are observed for the most at risk subgroup, or to stop for what is claimed to be safety reasons due to a random occurrence of an event during the first administration to a subgroup at the lowest dose.

While estimating different TDs for patients with certain baseline characteristics is a reasonable procedure to undertake, one cannot predict an individual patient's underlying tolerance to the investigational drug. While there may be some relationship between the drug being administered and perhaps the presence of a biomarker or another current medical condition, these may not be known or present at the onset of treatment. Therefore it is not possible to dose different patients at different doses to suit their underlying tolerability at the start of treatment. However, the ability to incorporate intra-patient dose adjustments, in order to ensure safe escalations that are suited to each patient, is desirable. The most effective way of incorporating intra-patient adjustments into the ICSDP is to allow adjustments between cycles based on accruing information in each cycle, but estimating the TD at the end of the trial to correspond just to the occurrence of DLTs in the entire population. Different patients will, in reality, require different TDs, and this may be indicated by the value of a tolerance marker. However, it may be appropriate to recommend a population average TD for further investigation, as dose-titration may be considered later on if a

significant difference in tolerance arises. There are different events or measurements that could be considered as a marker for tolerance, such as the occurrence of LGTs, some pharmacokinetic measure depicting the exposure of the drug, or a pharmacodynamic event which can be related to efficacy. The inclusion of LGTs is however one of the easiest to implement since all toxicities of all grades are recorded at every evaluation for each patient as a standard safety assessment, and so are readily attainable.

On including LGTs into the between cycle escalations for patients, it has been shown that the trial lengths can be decreased further, since each patient is being dosed at doses most suited to them so the dose corresponding to the required TTL is more precise. Escalating the doses between cohorts based on the estimated TDs associated with the occurrence of just DLTs (without LGTs) is very effective and requires no additional complexity in the procedure. Once the trials have stopped, the average of the population's individual TDs produces even better estimates than when analysing the data for just DLTs and obtaining a single TD. Furthermore, some idea can be gained on the effect of LGTs on the prediction of DLTs which can then be used in later phases of development to aid dosing decisions when LGTs occur.

9.2. Extensions and Further Work

While many possible scenarios have been investigated within the scope of each comparison, there are possible extensions and further comparisons that could have been considered for each chapter.

The inclusion of three cycles in the simulations of the ICSDP was based on the data from Postel-Vinay [1]. The Christie data [17] also only investigate for 3 cycles even when considering LGTs and looking at biomarker values, which suggests that 3 may

be sufficient. Also, if the suggestion of P(DLT) halving with successive cycles is appropriate, one would expect cycles after cycle 3 to contribute very few occurrences of DLTs. While the use of 3 cycles is justified in this thesis, some further investigation could be conducted to compare the use of the ICSDP with differing numbers of cycles to determine an optimal number of cycles to observe.

The scenarios investigated were based on the assumption that the probability of a patient's first DLT in each cycle was conditionally independent on P(DLT) in earlier cycles and consistently decreasing. It may be appropriate to consider the possibility that the probability of a patient's first DLT may increase with cycle due to an accumulation of dose. This could be investigated by including a cumulative dose term in the link function for the ICS model. However, there may be the possibility of an increased tolerance with time, so some compromise of decreasing P(DLT) with time, but increasing P(DLT) due to dose accumulation could be investigated. The cumulative dose term will be a time-changing covariate, so the methods in Chapter 8 would have to be applied to investigate this, and of course more parameters would require to be estimated. This may discourage the inclusion of further covariates which would then require more parameters still.

Prior information has been incorporated through the use of pseudo-data, which combines easily with the binary observations from the trial. Further investigation could be undertaken concerning the inclusion of prior information through the traditional method of meta-analytical priors, which use existing data to place a distribution on the parameters of the model which reflect current belief associated with the drug. This does however detract from the simplicity of the ICSDP and may be more complicated to implement.

The inclusion of baseline covariates could be extended to the inclusion of continuous covariates as discussed in Chapter 7. A TD to recommend for further investigation could be recommended as a function of the covariates such that a decision on dosing could be made based on the specific patient's characteristics at the start of the trial.

The results from Chapter 8 could be combined with the inclusion of baseline covariates to determine whether different doses are appropriate to administer at the start of treatment while allowing dose adjustments based on time changing covariates. Further investigation into the effect of time-changing covariates across baseline characteristics could also be investigated. For example the presence of characteristic specific markers, such as hormone levels which have been related to particular cancers, could fluctuate in response to the drug which could provide early indications of efficacy or toxicity. Those markers may have a different level of prevalence in different patients, females to males for example, so the effect on P(DLT) may be different for those patients, and therefore interaction terms between the time-changing covariates and the baseline characteristics may need to be included.

10. Appendices

Appendix 1:

Rearranging the complementary log-log link function in terms of the TD for s cycles.

The link function for the conditional probability is defined as:

$$\pi_{j,l} = 1 - \exp\left(-\exp\left(\gamma_l + \theta \log d_j\right)\right)$$

$$\log\left(-\log\left(1-\pi_{j,l}\right)\right) = \gamma_l + \theta \log d_j$$

$$p_j(c_s) = \pi_{j1} + \left(1-\pi_{j1}\right)\pi_{j2} + \ldots + \left(1-\pi_{j1}\right)\cdots\left(1-\pi_{j,s-1}\right)\pi_{js}.$$

The probability after s cycles redefined through the link function is:

$$\begin{aligned}
&p_j(c_s) = \\
&1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_j\right)\right) + \\
&\exp\left(-\exp\left(\gamma_1 + \theta \log d_j\right)\right)\left[1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_j\right)\right)\right] + \\
&\vdots \\
&+\exp\left(-\exp\left(\gamma_1 + \theta \log d_j\right)\right)\cdots\exp\left(-\exp\left(\gamma_{s-1} + \theta \log d_j\right)\right)\left[1-\exp\left(-\exp\left(\gamma_s + \theta \log d_j\right)\right)\right] \\
&= 1-\left\{\exp\left(-\exp\left(\gamma_1 + \theta \log d_j\right)\right)\cdots\exp\left(-\exp\left(\gamma_{s-1} + \theta \log d_j\right)\right)\exp\left(-\exp\left(\gamma_s + \theta \log d_j\right)\right)\right\} \\
&= 1-\exp\left\{d_j^{\theta}\left[-e^{\gamma_1}\cdots-e^{\gamma_{s-1}}-e^{\gamma_s}\right]\right\}.
\end{aligned}$$

The TD can be defined by rearranging the expression for the TTL as follows:

$$TTL = 1-\exp\left\{TD^{\theta}\left[-e^{\gamma_1}\cdots-e^{\gamma_{s-1}}-e^{\gamma_s}\right]\right\}$$

$$\log(TD) = \frac{1}{\theta}\left[\log\left(\frac{-\log(1-TTL)}{e^{\gamma_1}\cdots+e^{\gamma_{s-1}}+e^{\gamma_s}}\right)\right]$$

$$TD = \exp\left\{\frac{1}{\theta}\left[\log\left(\frac{-\log(1-TTL)}{e^{\gamma_1}\cdots+e^{\gamma_{s-1}}+e^{\gamma_s}}\right)\right]\right\}$$

$$TD = \left(\frac{-\log(1-TTL)}{e^{\gamma_1}\cdots+e^{\gamma_{s-1}}+e^{\gamma_s}}\right)^{1/\theta}.$$

Appendix 2:

Deriving the asymptotic variance for the complementary log-log link function for 3 cycles

Given $p_{j3} = \pi_{j3}\left(1-\pi_{j2}\right)\left(1-\pi_{j1}\right)$, the likelihood (L) is:

$$
\begin{aligned}
L &= \prod_{j=1}^{k}\prod_{l=1}^{3+1} p_{(j)l}^{\;t_{(j)l}} \\
&= \prod_{j=1}^{k}\prod_{l=1}^{3+1}\left[\left(1-\pi_{(j),l-(l-1)}\right)\left(1-\pi_{(j),l-(l-2)}\right)\ldots\left(1-\pi_{(j),l-1}\right)\pi_{(j)l}\right]^{t_{(j)l}} \\
&= \prod_{j=1}^{k}\pi_{(j)4}^{\;t_{(j)4}}\prod_{l=1}^{3}\pi_{(j)l}^{\;t_{(j)l}}\left(1-\pi_{(j)l}\right)^{q_{(j)l}} \\
&= \prod_{j=1}^{k}\prod_{l=1}^{3}\pi_{(j)l}^{\;t_{(j)l}}\left(1-\pi_{(j)l}\right)^{q_{(j)l}},
\end{aligned}
$$

where $q_{(j)l} = n_{(j)l} - t_{(j)l}$, the number of subjects on dose level (j) during cycle l who did not experience a DLT. L is then expressed through the link function as follows:

$$
\begin{aligned}
L = \prod_{j=1}^{k}&\left(1-\exp\left(-\exp\left(\gamma_1+\theta\log d_{(j)}\right)\right)\right)^{t_{(j)1}}\left(\exp\left(-\exp\left(\gamma_1+\theta\log d_{(j)}\right)\right)\right)^{q_{(j)1}} \\
&\left(1-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\right)^{t_{(j)2}}\left(\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\right)^{q_{(j)2}} \\
&\left(1-\exp\left(-\exp\left(\gamma_3+\theta\log d_{(j)}\right)\right)\right)^{t_{(j)3}}\left(\exp\left(-\exp\left(\gamma_3+\theta\log d_{(j)}\right)\right)\right)^{q_{(j)3}}.
\end{aligned}
$$

Taking the natural logarithm gives the expression for the log-likelihood.

$$
\begin{aligned}
\ell = \sum_{j=1}^{k}&t_{(j)1}\log\left(1-\exp\left(-\exp\left(\gamma_1+\theta\log d_{(j)}\right)\right)\right)-q_{(j)1}\left(\exp\left(\gamma_1+\theta\log d_{(j)}\right)\right) \\
&t_{(j)2}\log\left(1-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\right)-q_{(j)2}\left(\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right) \\
&t_{(j)3}\log\left(1-\exp\left(-\exp\left(\gamma_3+\theta\log d_{(j)}\right)\right)\right)-q_{(j)3}\left(\exp\left(\gamma_3+\theta\log d_{(j)}\right)\right)
\end{aligned}
$$

Differentiating with respect to each parameter:

$$\frac{\partial \ell}{\partial \gamma_1} = \sum_{j=1}^{k} \frac{t_{(j)1}\left(-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} - q_{(j)1} \exp\left(\gamma_1 + \theta \log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} -\frac{\log\left(1-\pi_{(j)1}\right)}{\pi_{(j)1}}\left\{t_{(j)1} - \pi_{(j)1}\left(t_{(j)1} + q_{(j)1}\right)\right\}$$

$$\frac{\partial \ell}{\partial \gamma_2} = \sum_{j=1}^{k} \frac{t_{(j)2}\left(-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)} - q_{(j)2} \exp\left(\gamma_2 + \theta \log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} -\frac{\log\left(1-\pi_{(j)2}\right)}{\pi_{(j)2}}\left\{t_{(j)2} - \pi_{(j)2}\left(t_{(j)2} + q_{(j)2}\right)\right\}$$

$$\frac{\partial \ell}{\partial \gamma_3} = \sum_{j=1}^{k} \frac{t_{(j)3}\left(-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)} - q_{(j)3} \exp\left(\gamma_3 + \theta \log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} -\frac{\log\left(1-\pi_{(j)3}\right)}{\pi_{(j)3}}\left\{t_{(j)3} - \pi_{(j)3}\left(t_{(j)3} + q_{(j)3}\right)\right\}$$

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^{k} \log d_{(j)} \left\{ \frac{t_{(j)1}\left(-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} - q_{(j)1} \exp\left(\gamma_1 + \theta \log d_{(j)}\right) \right\}$$

$$+ \log d_{(j)} \left\{ \frac{t_{(j)2}\left(-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)} - q_{(j)2} \exp\left(\gamma_2 + \theta \log d_{(j)}\right) \right\}$$

$$+ \log d_{(j)} \left\{ \frac{t_{(j)3}\left(-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)} - q_{(j)3} \exp\left(\gamma_3 + \theta \log d_{(j)}\right) \right\}$$

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^{k} \log d_{(j)} \left[-\frac{\log\left(1-\pi_{(j)1}\right)}{\pi_{(j)1}}\left\{t_{(j)1} - \pi_{(j)1}\left(t_{(j)1} + q_{(j)1}\right)\right\} \right.$$

$$\left. -\frac{\log\left(1-\pi_{(j)2}\right)}{\pi_{(j)2}}\left\{t_{(j)2} - \pi_{(j)2}\left(t_{(j)2} + q_{(j)2}\right)\right\} - \frac{\log\left(1-\pi_{(j)3}\right)}{\pi_{(j)3}}\left\{t_{(j)3} - \pi_{(j)3}\left(t_{(j)3} + q_{(j)3}\right)\right\} \right].$$

Differentiating again:

$$\frac{\partial^2 \ell}{\partial \gamma_1 \partial \gamma_2} = \frac{\partial^2 \ell}{\partial \gamma_1 \partial \gamma_3} = \frac{\partial^2 \ell}{\partial \gamma_3 \partial \gamma_2} = 0$$

$$\frac{\partial^2 \ell}{\partial \gamma_1^{\,2}} = \sum_{j=1}^{k} \frac{t_{(j)1}\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)}$$

$$+\frac{t_{(j)1}\exp\left(\gamma_1 + \theta \log d_{(j)}\right)}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} - n_{(j)1}\exp\left(\gamma_1 + \theta \log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} \frac{-t_{(j)1}\left(\log\left(1-\pi_{(j)1}\right)\right)^2 - t_{(j)1}\pi_{(j)1}\log\left(1-\pi_{(j)1}\right)\left\{1-\log\left(1-\pi_{(j)1}\right)\right\} + n_{(j)1}\pi_{(j)1}^{\,2}\log\left(1-\pi_{(j)1}\right)}{\pi_{(j)1}^{\,2}}$$

$$\frac{\partial^2 \ell}{\partial \gamma_2^{\,2}} = \sum_{j=1}^{k} \frac{t_{(j)2}\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)}$$

$$+\frac{t_{(j)1}\exp\left(\gamma_2 + \theta \log d_{(j)}\right)}{1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)} - n_{(j)2}\exp\left(\gamma_2 + \theta \log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} \frac{-t_{(j)2}\left(\log\left(1-\pi_{(j)2}\right)\right)^2 - t_{(j)2}\pi_{(j)2}\log\left(1-\pi_{(j)2}\right)\left\{1-\log\left(1-\pi_{(j)2}\right)\right\} + n_{(j)2}\pi_{(j)2}^{\,2}\log\left(1-\pi_{(j)2}\right)}{\pi_{(j)2}^{\,2}}$$

$$\frac{\partial^2 \ell}{\partial \gamma_3^{\,2}} = \sum_{j=1}^{k} \frac{t_{(j)3}\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)}$$

$$+\frac{t_{(j)3}\exp\left(\gamma_3 + \theta \log d_{(j)}\right)}{1-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)} - n_{(j)3}\exp\left(\gamma_3 + \theta \log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} \frac{-t_{(j)3}\left(\log\left(1-\pi_{(j)3}\right)\right)^2 - t_{(j)3}\pi_{(j)3}\log\left(1-\pi_{(j)3}\right)\left\{1-\log\left(1-\pi_{(j)3}\right)\right\} + n_{(j)3}\pi_{(j)3}^{\,2}\log\left(1-\pi_{(j)3}\right)}{\pi_{(j)3}^{\,2}}$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{j=1}^{k} \log d_{(j)} \times$$

$$\left\{ \frac{t_{(j)1}\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\log d_{(j)}}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} \right.$$

$$+ \frac{t_{(j)1}\left(\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\log d_{(j)}}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} - n_{(j)1}\left(\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\log d_{(j)}$$

$$+ \frac{t_{(j)2}\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\log d_{(j)}}{1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)}$$

$$+ \frac{t_{(j)2}\left(\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\log d_{(j)}}{1-\exp\left(-\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)} - n_{(j)2}\left(\exp\left(\gamma_2 + \theta \log d_{(j)}\right)\right)\log d_{(j)}$$

$$+ \frac{t_{(j)3}\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\log d_{j}}{1-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)}$$

$$\left. + \frac{t_{(j)3}\left(\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\log d_{(j)}}{1-\exp\left(-\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)} - n_{(j)3}\left(\exp\left(\gamma_3 + \theta \log d_{(j)}\right)\right)\log d_{(j)} \right\}$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{j=1}^{k} \left(\log d_{(j)}\right)^2$$

$$\left\{ \frac{-t_{(j)1}\left(\log\left(1-\pi_{(j)1}\right)\right)^2 - t_{(j)1}\pi_{(j)1}\log\left(1-\pi_{(j)1}\right)\left\{1-\log\left(1-\pi_{(j)1}\right)\right\} + n_{(j)1}{\pi_{(j)1}}^2 \log\left(1-\pi_{(j)1}\right)}{{\pi_{(j)1}}^2} \right.$$

$$+ \frac{-t_{(j)2}\left(\log\left(1-\pi_{(j)2}\right)\right)^2 - t_{(j)2}\pi_{(j)2}\log\left(1-\pi_{(j)2}\right)\left\{1-\log\left(1-\pi_{(j)2}\right)\right\} + n_{(j)2}{\pi_{(j)2}}^2 \log\left(1-\pi_{(j)2}\right)}{{\pi_{j2}}^2}$$

$$\left. \frac{-t_{(j)3}\left(\log\left(1-\pi_{(j)3}\right)\right)^2 - t_{(j)3}\pi_{(j)3}\log\left(1-\pi_{(j)3}\right)\left\{1-\log\left(1-\pi_{(j)3}\right)\right\} + n_{(j)3}{\pi_{(j)3}}^2 \log\left(1-\pi_{(j)3}\right)}{{\pi_{(j)3}}^2} \right\}$$

$$\frac{\partial^2 \ell}{\partial \gamma_1 \partial \theta} = \sum_{j=1}^{k} \log d_{(j)} \times$$

$$\left\{ \frac{t_{(j)1}\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} \right.$$

$$\left. + \frac{t_{(j)1}\left(\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right)} - n_{(j)1}\left(\exp\left(\gamma_1 + \theta \log d_{(j)}\right)\right) \right\}$$

$$\frac{\partial^2 \ell}{\partial \gamma_1 \partial \theta} = \sum_{j=1}^{k} \log d_{(j)}$$

$$\left\{ \frac{-t_{(j)1}\left(\log\left(1-\pi_{(j)1}\right)\right)^2 - t_{(j)1}\pi_{(j)1}\log\left(1-\pi_{(j)1}\right)\left\{1-\log\left(1-\pi_{(j)1}\right)\right\} + n_{(j)1}\pi_{(j)1}{}^2\log\left(1-\pi_{(j)1}\right)}{\pi_{(j)1}{}^2} \right\}$$

$$\frac{\partial^2 \ell}{\partial \gamma_2 \partial \theta} = \sum_{j=1}^{k} \log d_{(j)} \times$$

$$\left\{ \frac{t_{(j)2}\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)} \right.$$

$$\left. + \frac{t_{(j)2}\left(\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)} - n_{(j)2}\left(\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right) \right\}$$

$$\frac{\partial^2 \ell}{\partial \gamma_2 \partial \theta} = \sum_{j=1}^{k} \log d_{(j)}$$

$$\left\{ \frac{-t_{(j)2}\left(\log\left(1-\pi_{(j)2}\right)\right)^2 - t_{(j)2}\pi_{(j)2}\log\left(1-\pi_{(j)2}\right)\left\{1-\log\left(1-\pi_{(j)2}\right)\right\} + n_{(j)2}\pi_{(j)2}{}^2\log\left(1-\pi_{(j)2}\right)}{\pi_{(j)2}{}^2} \right\}$$

$$\frac{\partial^2 \ell}{\partial \gamma_3 \partial \theta} = \sum_{j=1}^{k} \log d_{(j)} \times$$

$$\left\{ \frac{t_{(j)2}\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\left(-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)} \right.$$

$$\left. + \frac{t_{(j)2}\left(\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)}{1-\exp\left(-\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right)} - n_{(j)2}\left(\exp\left(\gamma_2+\theta\log d_{(j)}\right)\right) \right\}$$

$$\frac{\partial^2 \ell}{\partial \gamma_3 \partial \theta} = \sum_{j=1}^{k} \log d_{(j)}$$

$$\left\{ \frac{-t_{(j)3}\left(\log\left(1-\pi_{(j)3}\right)\right)^2 - t_{(j)3}\pi_{(j)3}\log\left(1-\pi_{(j)3}\right)\left\{1-\log\left(1-\pi_{(j)3}\right)\right\} + n_{(j)3}\pi_{(j)3}{}^2\log\left(1-\pi_{(j)3}\right)}{\pi_{(j)3}{}^2} \right\}.$$

Letting;

$$R_{(j)1} = \frac{t_{(j)1}\pi_{(j)1}{}^2\log\left(1-\pi_{(j)1}\right) - t_{(j)1}\left(\log\left(1-\pi_{(j)1}\right)\right)^2 - t_{(j)1}\pi_{(j)1}\log\left(1-\pi_{(j)1}\right)\left\{1-\log\left(1-\pi_{(j)1}\right)\right\}}{\pi_{(j)1}{}^2}$$

$$R_{(j)2}=\frac{t_{(j)2}\pi_{(j)2}{}^{2}\log\left(1-\pi_{(j)2}\right)-t_{(j)2}\left(\log\left(1-\pi_{(j)2}\right)\right)^{2}-t_{(j)2}\pi_{(j)2}\log\left(1-\pi_{(j)2}\right)\left\{1-\log\left(1-\pi_{(j)2}\right)\right\}}{\pi_{(j)2}{}^{2}}$$

$$R_{(j)3}=\frac{t_{(j)3}\pi_{(j)3}{}^{2}\log\left(1-\pi_{(j)3}\right)-t_{(j)3}\left(\log\left(1-\pi_{(j)3}\right)\right)^{2}-t_{(j)3}\pi_{(j)3}\log\left(1-\pi_{(j)3}\right)\left\{1-\log\left(1-\pi_{(j)3}\right)\right\}}{\pi_{(j)3}{}^{2}}.$$

The 2nd derivative Matrix is then defined as;

$$\begin{pmatrix}\sum_{j=1}^{k}R_{(j)1} & 0 & 0 & \sum_{j=1}^{k}R_{(j)1}(\log d_{(j)})\\ 0 & \sum_{j=1}^{k}R_{(j)2} & 0 & \sum_{j=1}^{k}R_{(j)2}(\log d_{(j)})\\ 0 & 0 & \sum_{j=1}^{k}R_{(j)3} & \sum_{j=1}^{k}R_{(j)3}(\log d_{(j)})\\ \sum_{j=1}^{k}R_{(j)1}(\log d_{j}) & \sum_{j=1}^{k}R_{(j)2}(\log d_{j}) & \sum_{j=1}^{k}R_{(j)3}(\log d_{j}) & \sum_{j=1}^{k}(R_{(j)1}+R_{(j)2}+R_{(j)3})(\log d_{(j)})\end{pmatrix}$$

The Observed Information Matrix is the negative of the 2nd derivative matrix;

$I_{O}(\gamma_{1},\gamma_{2},\gamma_{3},\theta)=$

$$\begin{pmatrix}-\sum_{j=1}^{k}R_{(j)1} & 0 & 0 & -\sum_{j=1}^{k}R_{(j)1}(\log d_{(j)})\\ 0 & -\sum_{j=1}^{k}R_{(j)2} & 0 & -\sum_{j=1}^{k}R_{(j)2}(\log d_{(j)})\\ 0 & 0 & -\sum_{j=1}^{k}R_{(j)3} & -\sum_{j=1}^{k}R_{(j)3}(\log d_{(j)})\\ -\sum_{j=1}^{k}R_{(j)1}(\log d_{j}) & -\sum_{j=1}^{k}R_{(j)2}(\log d_{j}) & -\sum_{j=1}^{k}R_{(j)3}(\log d_{j}) & -\sum_{j=1}^{k}(R_{(j)1}+R_{(j)2}+R_{(j)3})(\log d_{(j)})\end{pmatrix}$$

The Expected Information Matrix: $I_{E}(\gamma_{1},\gamma_{2},\gamma_{3},\theta)=E\left[I_{O}(\gamma_{1},\gamma_{2},\gamma_{3},\theta)\right]$

The determinant of this matrix required to invert this matrix is given by the following expression;

$$\det = \sum_{j=1}^{k} R_{(j)1} \sum_{j=1}^{k} R_{(j)2} \sum_{j=1}^{k} R_{(j)3} \sum_{j=1}^{k} \left(R_{(j)1} + R_{(j)2} + R_{(j)3}\right) \left(\log d_{(j)}\right)^2$$
$$- \left(\sum_{j=1}^{k} R_{(j)1} \log d_{(j)}\right)^2 \sum_{j=1}^{k} R_{(j)2} \sum_{j=1}^{k} R_{(j)3} - \left(\sum_{j=1}^{k} R_{(j)2} \log d_{(j)}\right)^2 \sum_{j=1}^{k} R_{(j)1} \sum_{j=1}^{k} R_{(j)3}$$
$$- \left(\sum_{j=1}^{k} R_{(j)3} \log d_{(j)}\right)^2 \sum_{j=1}^{k} R_{(j)1} \sum_{j=1}^{k} R_{(j)2}$$

The inverted Expected Information Matrix is shown in Appendix 3.

Replace $d_{(j)}$ with TD and rearrange the complementary log-log link function in terms of $\log(TD)$ as shown in Appendix 1.

The first derivatives of $\log(TD)$ with respect to each of the parameters are;

$$\frac{\partial \log(TD)}{\partial \gamma_1} = -\frac{1}{\theta} \frac{1}{\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)} e^{\gamma_1} = -\frac{e^{\gamma_1}}{\theta\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)}$$
$$\frac{\partial \log(TD)}{\partial \gamma_2} = -\frac{1}{\theta} \frac{1}{\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)} e^{\gamma_2} = -\frac{e^{\gamma_2}}{\theta\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)}$$

$$\frac{\partial \log(TD)}{\partial \gamma_3} = -\frac{1}{\theta} \frac{1}{\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)} e^{\gamma_3} = -\frac{e^{\gamma_3}}{\theta\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)}$$

$$\frac{\partial \log(TD)}{\partial \theta} = -\frac{1}{\theta^2} \log\left(-\log\left(1 - p_j(c_3)\right)\right) + \frac{1}{\theta^2} \log\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)$$
$$= -\frac{1}{\theta}\left\{\frac{1}{\theta} \log\left(-\log\left(1 - p_j(c_3)\right)\right) - \frac{1}{\theta} \log\left(e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}\right)\right\}$$
$$= -\frac{1}{\theta} \log(TD)$$

The first derivative vector of $\log(TD)$:

$$\nabla\left(\log(TD)\right)^T = -\frac{1}{\theta}\left(\frac{e^{\gamma_1}}{e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}}, \frac{e^{\gamma_2}}{e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}}, \frac{e^{\gamma_3}}{e^{\gamma_1} + e^{\gamma_2} + e^{\gamma_3}}, \log(TD)\right)$$

The asymptotic variance of $\log(TD)$ is given by;

$$\nabla\left(\log(TD)\right)^T I_E^{-1}(\gamma_1, \gamma_2, \gamma_3, \theta) \nabla\left(\log(TD)\right)$$

The asymptotic variance is then defined as;

$$\frac{1}{\theta^2 \det}\times\left\{\left(\frac{e^{\gamma_1}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)^2\left(\frac{\det+\left(-\sum_{j=1}^{k}R_{(j)2}\right)\left(-\sum_{j=1}^{k}R_{(j)3}\right)\left(-\sum_{j=1}^{k}R_{(j)1}\log d_j\right)^2}{\left(-\sum_{j=1}^{k}R_{(j)1}\right)}\right)\right.$$

$$-2\left(\frac{e^{\gamma_1}e^{\gamma_2}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)-2\left(\frac{e^{\gamma_1}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)+2\left(\frac{e^{\gamma_1}\log(TD)}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)$$

$$+\left(\frac{e^{\gamma_2}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)\left(\frac{\det+\left(-\sum_{j=1}^{k}R_{(j)1}\right)\left(-\sum_{j=1}^{k}R_{(j)3}\right)\left(-\sum_{j=1}^{k}R_{(j)2}\log d_j\right)^2}{\left(-\sum_{j=1}^{k}R_{(j)2}\right)}\right)$$

$$-2\left(\frac{e^{\gamma_2}e^{\gamma_1}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)-2\left(\frac{e^{\gamma_2}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)+2\left(\frac{e^{\gamma_2}\log(TD)}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)$$

$$+\left(\frac{e^{\gamma_3}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)\left(\frac{\det+\left(-\sum_{j=1}^{k}R_{(j)1}\right)\left(-\sum_{j=1}^{k}R_{(j)2}\right)\left(-\sum_{j=1}^{k}R_{(j)3}\log d_j\right)^2}{\left(-\sum_{j=1}^{k}R_{(j)3}\right)}\right)$$

$$-2\left(\frac{e^{\gamma_1}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)-2\left(\frac{e^{\gamma_2}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)+2\left(\frac{e^{\gamma_3}\log(TD)}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)$$

$$\left.-\left(\log(TD)\right)^2\sum_{j=1}^{k}R_{(j)1}\sum_{j=1}^{k}R_{(j)2}\sum_{j=1}^{k}R_{(j)3}\right\}.$$

Appendix 3:
Inverted Information Matrix

$$I_E^{-1}(\gamma_1,\gamma_2,\gamma_3,\theta) = 1/\det \times$$

$$\begin{pmatrix}
\dfrac{\det+\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right)^2}{\left(-\sum_{j=1}^{k} R_{(j)1}\right)} & \left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\log d_{(j)}\right) & \left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\log d_{(j)}\right) & \left(-\sum_{i=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right) \\
\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\log d_{(j)}\right) & \dfrac{\det+\left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\log d_{(j)}\right)^2}{\left(-\sum_{j=1}^{k} R_{(j)1}\right)} & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R(j)_2\log d_{(j)}\right)\left(-\sum_{j=1}^{k} R(j)_3\log d_{(j)}\right) & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\log d_{(j)}\right) \\
\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\log d_{(j)}\right) & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\log d_{(j)}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\log d_{(j)}\right) & \dfrac{\det+\left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\log d_{(j)}\right)^2}{\left(-\sum_{j=1}^{k} R_{(j)3}\right)} & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\log d_{(j)}\right) \\
\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)1}\log d_{(j)}\right) & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\log d_{(j)}\right) & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\log d_{(j)}\right) & \left(-\sum_{j=1}^{k} R_{(j)1}\right)\left(-\sum_{j=1}^{k} R_{(j)2}\right)\left(-\sum_{j=1}^{k} R_{(j)3}\right)
\end{pmatrix}$$

Appendix 4:
Extending the asymptotic variance for the complementary log-log link function to s cycles

As in Appendix 2, let:

$$R_{(j)l} = \frac{n_{(j)l}\pi_{(j)l}^{\ 2}\log\left(1-\pi_{(j)l}\right) - t_{(j)l}\left(\log\left(1-\pi_{(j)l}\right)\right)^2 - t_{(j)l}\pi_{(j)l}\log\left(1-\pi_{(j)}\right)\left\{1-\log\left(1-\pi_{(j)l}\right)\right\}}{\pi_{(j)l}^{\ 2}}$$

For cycle l, $l = 1,...,s$,

$$I_O(\gamma_1,...,\gamma_{s-1},\gamma_s,\theta) = I_E(\gamma_1,...,\gamma_{s-1},\gamma_s,\theta) =$$

$$\begin{bmatrix} -\sum_{j=1}^{k} R_{(j)1} & 0 & \cdots & 0 & -\sum_{j=1}^{k} R_{(j)1}\log d_{(j)} \\ 0 & -\sum_{j=1}^{k} R_{(j)2} & \cdots & 0 & -\sum_{j=1}^{k} R_{(j)2}\log d_{(j)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\sum_{j=1}^{k} R_{(j)s} & -\sum_{j=1}^{k} R_{(j)s}\log d_{(j)} \\ -\sum_{j=1}^{k} R_{(j)1}\log d_{(j)} & -\sum_{j=1}^{k} R_{(j)2}\log d_{(j)} & \cdots & -\sum_{j=1}^{k} R_{(j)s}\log d_{(j)} & -\sum_{j=1}^{k}\left(\log d_{(j)}\right)^2 \sum_{l=1}^{s} R_l \end{bmatrix}$$

$\log(TD)$ is defined as:

$$\log(TD) = \frac{1}{\theta}\log\left(-\log\left(1-TTL\right)\right) - \frac{1}{\theta}\log\left(\sum_{l=1}^{s} e^{\gamma_l}\right).$$

Differentiating with respect to each parameter gives:

$$\frac{\partial\log(TD)}{\partial\gamma_l} = -\frac{e^{\gamma_1}}{\theta\sum_{l=1}^{s} e^{\gamma_l}}, \text{ for } l = 1,...,s$$

$$\begin{aligned} \frac{\partial\log(TD)}{\partial\theta} &= -\frac{1}{\theta^2}\log\left(-\log\left(1-p_j\left(c_s\right)\right)\right) + \frac{1}{\theta^2}\log\left(e^{\gamma_1}+e^{\gamma_2}+...+e^{\gamma_s}\right) \\ &= -\frac{1}{\theta}\left\{\frac{1}{\theta}\log\left(-\log\left(1-p_j\left(c_s\right)\right)\right) - \frac{1}{\theta}\log\left(e^{\gamma_1}+e^{\gamma_2}+...+e^{\gamma_s}\right)\right\} \\ &= -\frac{1}{\theta}\log(TD). \end{aligned}$$

Therefore:

$$\nabla\left(\log\left(TD\right)\right)^{T} = -\frac{1}{\theta}\left(\frac{e^{\gamma_1}}{\sum_{l=1}^{s} e^{\gamma_l}}, \frac{e^{\gamma_2}}{\sum_{l=1}^{s} e^{\gamma_l}}, \ldots, \frac{e^{\gamma_s}}{\sum_{l=1}^{s} e^{\gamma_l}}, \log\left(TD\right)\right).$$

The determinant required to invert the information matrix is defined as:

$$\det = \left[\prod_{l=1}^{s}\left(\sum_{j=1}^{k} R_{(j)l}\right)\right]\sum_{j=1}^{k}\left\{\left(\log d_{(j)}\right)^{2}\left(\sum_{l=1}^{s} R_{(j)l}\right)\right\} - \sum_{l=1}^{s}\left[\left(\sum_{j=1}^{k} R_{(j)l}\log d_{(j)}\right)^{2}\prod_{t\in(1,s)\setminus l}\left(\sum_{j=1}^{k} R_{(j)t}\right)\right].$$

The asymptotic variance is then derived:

$$\operatorname{var}\left(\log\left(TD\right)\right) =$$

$$\frac{1}{\theta^{2}\det}\left\{\sum_{l=1}^{s}\left(\frac{e^{\gamma_l}}{\sum_{l=1}^{s} e^{\gamma_l}}\right)\left\{\left(\frac{e^{\gamma_l}}{\sum_{l=1}^{s} e^{\gamma_l}}\right)\left(\frac{\det + \left(\sum_{j=1}^{k} R_{(j)l}\log d_{(j)}\right)^{2}\prod_{t\in(1,s)\setminus l}\left(\prod_{j=1}^{k} R_{(j)t}\right)}{-\sum_{j=1}^{k} R_{(j)l}}\right)\right.\right.$$

$$-2\left[\sum_{t\in(1,s)\setminus l}\left(\frac{e^{\gamma_t}}{\sum_{l=1}^{s} e^{\gamma_l}}\right)\left(\sum_{j=1}^{k} R_{(j)l}\log d_{(j)}\sum_{j=1}^{k} R_{(j)t}\log d_{(j)}\prod_{u\in(1,s)\setminus l,t}\left(\sum_{j=1}^{k} R_{(j)u}\right)\right)\right]$$

$$\left.\left.+2\log\left(TD\right)\left[\sum_{j=1}^{k} R_{(j)l}\log d_{(j)}\left(\prod_{t\in(1,s)\setminus l}\left(\sum_{j=1}^{k} R_{(j)t}\right)\right)\right]\right\} - \left(\log\left(TD\right)\right)^{2}\left[\prod_{l=1}^{s}\left(\sum_{j=1}^{k} R_{(j)l}\right)\right]\right\}$$

Appendix 5:

Deriving the asymptotic variance for the complementary log-log link function for 3 cycles with 2 covariates

The probability of DLT during a given cycle l on dose level (j), for a patient with covariates a and g:

$$p_{(j),a,g,l} = \pi_{(j),a,g,l}\left(1-\pi_{(j),a,g,l-1}\right)...\left(1-\pi_{(j),a,g,1}\right).$$

The overall probability of DLT after 3 cycles on dose level (j).

$$\begin{aligned} p_{(j),a,g}(c_3) &= \pi_{(j),a,g1} + \pi_{(j),a,g2}\left(1-\pi_{(j),a,g1}\right) + \pi_{(j),a,g3}\left(1-\pi_{(j),a,g2}\right)\left(1-\pi_{(j),a,g1}\right) \\ &= 1-\exp\Big[-\exp\left(\gamma_1+\xi a+\nu g+\theta\log d_{(j)}\right)-\exp\left(\gamma_2+\xi a+\nu g+\theta\log d_{(j)}\right) \\ &\quad -\exp\left(\gamma_3+\xi a+\nu g+\theta\log d_{(j)}\right)\Big] \end{aligned}$$

$$\begin{aligned} \log\left(1-p_{(j),a,g}(c_3)\right) &= \exp(\xi a)\exp(\nu g)\exp\left(\theta\log d_{(j)}\right) \\ &\quad \left[-\exp(\gamma_1)-\exp(\gamma_2)-\exp(\gamma_3)\right] \end{aligned}$$

The dose corresponding to a given probability is defined as:

$$\exp\left[\frac{\log\left(\dfrac{\log\left(1-p_{(j),a,g}(c_3)\right)}{e^{\xi a}e^{\nu g}\left[-e^{\gamma_1}-e^{\gamma_2}-e^{\gamma_3}\right]}\right)}{\theta}\right] = d_{(j)}$$

$$\frac{\log\left(\dfrac{\log\left(1-p_{(j),a,g}(c_3)\right)}{e^{\xi a}e^{\nu g}\left[-e^{\gamma_1}-e^{\gamma_2}-e^{\gamma_3}\right]}\right)}{\theta} = \log\left(d_{(j)}\right).$$

The likelihood (L) is defined as:

$$\begin{aligned} L &= \prod_{j=1}^{k}\prod_{l=1}^{3} p_{(j)a,g,l}^{\;t_{(j)a,g,l}} \\ &= \prod_{j=1}^{k}\prod_{l=1}^{3} \pi_{(j)a,g,l}^{\;t_{(j)a,g,l}}\left(1-\pi_{(j)a,g,l}\right)^{q_{(j)a,g,l}} \end{aligned}$$

$$\begin{aligned} L = \prod_{j=1}^{k}\prod_{l=1}^{3} &\left(1-\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)\right)^{t_{(j)a,g,l}} \\ &\left(\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)\right)^{q_{(j)a,g,l}} \end{aligned}$$

The log-likelihood is defined as:

$$\ell = \sum_{j=1}^{k}\sum_{l=1}^{3} t_{(j)a,g,l} \log\left(1-\exp\left(-\exp\left(\gamma_1+\theta\log d_{(j)}\right)\right)\right) - q_{(j)a,g,l}\exp\left(\gamma_1+\theta\log d_{(j)}\right)$$

For cycle $l=1,2,3$, the derivatives with respect to each parameter are:

$$\frac{\partial\ell}{\partial\gamma_l} = \sum_{j=1}^{k} \frac{t_{(j)a,g,l}\left(-\exp\left(-\exp\left(\gamma_l+\theta\log d_{(j)}\right)\right)\right)}{1-\exp\left(-\exp\left(\gamma_l+\theta\log d_{(j)}\right)\right)} - q_{(j)a,g,l}\exp\left(\gamma_l+\theta\log d_{(j)}\right)$$

$$= \sum_{j=1}^{k} \frac{t_{(j)a,g,l}\left[-\left(1-\pi_{(j)a,g,l}\right)\log\left(1-\pi_{(j)a,g,l}\right)\right]}{\pi_{(j)a,g,l}} - q_{(j)a,g,l}\left(-\log\left(1-\pi_{(j)a,g,l}\right)\right)$$

$$= \sum_{j=1}^{k} -\frac{t_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}} + n_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)$$

$$= \sum_{j=1}^{k} \frac{t_{(j)a,g,l}\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)}{1-\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)} + n_{(j)a,g,l}\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)$$

$$\frac{\partial\ell}{\partial\xi} =$$

$$\sum_{j=1}^{k}\sum_{l=1}^{3} \frac{t_{(j)a,g,l}\left(-\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)a}{1-\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)}$$

$$-q_{(j)a,g,l}\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)a$$

$$\frac{\partial\ell}{\partial\xi} = \sum_{j=1}^{k}\sum_{l=1}^{3} -a\frac{\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}}\left[t_{(j)a,g,l} - \pi_{(j)a,g,l}\left(t_{(j)a,g,l}+q_{(j)a,g,l}\right)\right]$$

$$= \sum_{j=1}^{k}\sum_{l=1}^{3} -a\frac{t_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}} + an_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)$$

$$= \sum_{j=1}^{k}\sum_{l=1}^{3} a\frac{t_{(j)a,g,l}\exp\left(\gamma_1+\xi a+\nu g+\theta\log d_{(j)}\right)}{1-\exp\left(-\exp\left(\gamma_1+\xi a+\nu g+\theta\log d_{(j)}\right)\right)} - an_{(j)a,g,l}\exp\left(\gamma_1+\xi a+\nu g+\theta\log d_{(j)}\right)$$

$$\frac{\partial\ell}{\partial\nu} =$$

$$\sum_{j=1}^{k}\sum_{l=1}^{3} \frac{t_{(j)a,g,l}\left(-\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)\right)\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)g}{1-\exp\left(-\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)\right)}$$

$$-q_{(j)a,g,l}\exp\left(\gamma_l+\xi a+\nu g+\theta\log d_{(j)}\right)g$$

$$\frac{\partial\ell}{\partial\nu} = \sum_{j=1}^{k}\sum_{l=1}^{3} -g\frac{\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}}\left[t_{(j)a,g,l} - \pi_{(j)a,g,l}\left(t_{(j)a,g,l}+q_{(j)a,g,l}\right)\right]$$

$$\frac{\partial\ell}{\partial\theta} = \sum_{j=1}^{k}\sum_{l=1}^{3} \log d_{(j)}\left\{-\frac{\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}}\left[t_{(j)a,g,l} - \pi_{(j)a,g,l}\left(t_{(j)a,g,l}+q_{(j)a,g,l}\right)\right]\right\}.$$

For $l = 1,2,3$, the second derivatives are:

$$\frac{\partial^2 \ell}{\partial \gamma_l^{\,2}} = \sum_{j=1}^{k} \left\{ - \frac{t_{(j)a,g,l} e^{\gamma_l + \xi a + \nu g + \theta \log d_{(j)}} \left(-e^{-\exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right)} \right) \left(-e^{\gamma_l + \xi a + \nu g + \theta \log d_{(j)}} \right)}{\left[1 - e^{\gamma_l + \xi a + \nu g + \theta \log d_{(j)}} \right]^2} \right.$$

$$\left. + \frac{t_{(j)a,g,l} e^{\gamma_l + \xi a + \nu g + \theta \log d_{(j)}}}{1 - e^{-\exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right)}} + n_{(j)a,g,l} e^{\gamma_l + \xi a + \nu g + \theta \log d_{(j)}} \right\}$$

$$= \sum_{j=1}^{k} \frac{-t_{(j)a,g,l} \left(\log\left(1 - \pi_{(j)a,g,l}\right) \right)^2 - t_{(j)a,g,l} \pi_{(j)a,g,l} \log\left(1 - \pi_{(j)a,g,l}\right) \left\{ 1 - \log\left(1 - \pi_{(j)a,g,l}\right) \right\}}{\pi_{(j)a,g,l}^{\,2}}$$

$$+ \frac{n_{(j)a,g,l} \pi_{(j)a,g,l}^{\,2} \log\left(1 - \pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\,2}}$$

$$\frac{\partial^2 \ell}{\partial \gamma_1 \partial \gamma_2} = \frac{\partial^2 \ell}{\partial \gamma_1 \partial \gamma_3} = \frac{\partial^2 \ell}{\partial \gamma_3 \partial \gamma_2} = 0$$

$$\frac{\partial^2 \ell}{\partial \xi^2}$$

$$= \sum_{j=1}^{k} \sum_{l=1}^{3} a \frac{t_{(j)a,g,l} e^{\gamma_l + \xi a + \nu g + \theta \log d_{(j)}} \left(-e^{-\exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right)} \right) a \left(-\exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right) \right)}{-\left[1 - e^{-\exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right)} \right]^2}$$

$$+ a \frac{t_{(j)a,g,l} \exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right) a}{1 - e^{-\exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right)}} - a n_{(j)a,g,l} \exp\left(\gamma_l + \xi a + \nu g + \theta \log d_{(j)}\right) a$$

$$= \sum_{j=1}^{k} \sum_{l=1}^{3} - \frac{t_{(j)a,g,l} \left(\log\left(1 - \pi_{(j)a,g,l}\right) \right) \left(1 - \pi_{(j)a,g,l}\right) \log\left(1 - \pi_{(j)a,g,l}\right) a^2}{-\pi_{(j)a,g,l}^{\,2}}$$

$$+ \frac{t_{(j)a,g,l} \left(-\log\left(1 - \pi_{(j)a,g,l}\right) \right) a^2 \pi_{(j)a,g,l}}{\pi_{(j)a,g,l}^{\,2}} - \frac{n_{(j)a,g,l} \left(-\log\left(1 - \pi_{(j)a,g,l}\right) \right) a^2 \pi_{(j)a,g,l}^{\,2}}{\pi_{(j)a,g,l}^{\,2}}$$

$$= \sum_{j=1}^{k} \sum_{l=1}^{3} \frac{-t_{(j)a,g,l} a^2 \left(\log\left(1 - \pi_{(j)a,g,l}\right) \right)^2 - t_{(j)a,g,l} a^2 \pi_{(j)a,g,l} \log\left(1 - \pi_{(j)a,g,l}\right) \left\{ 1 - \log\left(1 - \pi_{(j)a,g,l}\right) \right\}}{\pi_{(j)a,g,l}^{\,2}}$$

$$+ \frac{n_{(j)a,g,l} a^2 \pi_{(j)a,g,l}^{\,2} \log\left(1 - \pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\,2}}$$

$$\frac{\partial^2 \ell}{\partial \nu^2} =$$

$$= \sum_{j=1}^{k}\sum_{l=1}^{3} g \frac{t_{(j)a,g,l} e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}} \left(-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}\right) g \left(-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)\right)}{-\left[1-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}\right]^2}$$

$$+ g \frac{t_{(j)a,g,l} \exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right) g}{1-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}} - g n_{(j)a,g,l} \exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right) g$$

$$= \sum_{j=1}^{k}\sum_{l=1}^{3} - \frac{t_{(j)a,g,l} \left(\log\left(1-\pi_{(j)a,g,l}\right)\right)\left(1-\pi_{(j)a,g,l}\right)\log\left(1-\pi_{(j)a,g,l}\right) g^2}{-\pi_{(j)a,g,l}^{\ 2}}$$

$$+ \frac{t_{(j)a,g,l} \left(-\log\left(1-\pi_{(j)a,g,l}\right)\right) g^2 \pi_{(j)a,g,l}}{\pi_{(j)a,g,l}^{\ 2}} - \frac{n_{(j)a,g,l} \left(-\log\left(1-\pi_{(j)a,g,l}\right)\right) g^2 \pi_{(j)a,g,l}^{\ 2}}{\pi_{(j)a,g,l}^{\ 2}}$$

$$= \sum_{j=1}^{k}\sum_{l=1}^{3} \frac{-t_{(j)a,g,l} g^2 \left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l} g^2 \pi_{(j)a,g,l} \log\left(1-\pi_{(j)a,g,l}\right)\left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}^{\ 2}}$$

$$+ \frac{n_{(j)a,g,l} g^2 \pi_{(j)a,g,l}^{\ 2} \log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\ 2}}$$

For $l = 1,2,3$

$$\frac{\partial^2 \ell}{\partial \gamma_l \partial \xi} = \sum_{j=1}^{k} - \frac{t_{(j)a,g,l} e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}} \left(-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}\right)\left(-e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}}\right) a}{\left[1-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}\right]^2}$$

$$+ \frac{t_{(j)a,g,l} e^{e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}}} a}{1-e^{-\exp\left(e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}}\right)}} + n_{(j)a,g,l} e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}} a$$

$$= \sum_{j=1}^{k} \frac{-t_{(j)a,g,l} a \left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l} \pi_{(j)a,g,l} \log\left(1-\pi_{(j)a,g,l}\right) a \left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}^{\ 2}}$$

$$+ \frac{n_{(j)a,g,l} \pi_{(j)a,g,l}^{\ 2} a \log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\ 2}}$$

$$\frac{\partial^2 \ell}{\partial \gamma_l \partial \nu} = \sum_{j=1}^{k} - \frac{t_{(j)a,g,l} e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}} \left(-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}\right)\left(-e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}}\right) g}{\left[1-e^{-\exp\left(\gamma_l+\xi a+\nu g+\theta \log d_{(j)}\right)}\right]^2}$$

$$+ \frac{t_{(j)a,g,l} e^{e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}}} g}{1-e^{-\exp\left(e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}}\right)}} + n_{(j)a,g,l} e^{\gamma_l+\xi a+\nu g+\theta \log d_{(j)}} g$$

$$= \sum_{j=1}^{k} \frac{-t_{(j)a,g,l} g \left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l}\pi_{(j)a,g,l} \log\left(1-\pi_{(j)a,g,l}\right) g \left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}^{\ 2}}$$

$$+ \frac{n_{(j)a,g,l}\pi_{(j)a,g,l}^{\ 2} g \log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\ 2}}$$

$$\frac{\partial^2 \ell}{\partial \xi \partial \nu} =$$

$$\sum_{j=1}^{k}\sum_{l=1}^{3} ag \left\{ \frac{t_{(j)a,g,l}\left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)\pi_{(j)a,g,l}\left(1-\log\left(1-\pi_{(j)a,g,l}\right)\right)}{\pi_{(j)a,g,l}^{\ 2}} \right.$$

$$\left. + \frac{n_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)\pi_{(j)a,g,l}^{\ 2}}{\pi_{(j)a,g,l}^{\ 2}} \right\}$$

For $l = 1,2,3$

$$\frac{\partial^2 \ell}{\partial \gamma_l \partial \theta} =$$

$$\sum_{j=1}^{k} \log d_{(j)} \left\{ \frac{-t_{(j)a,g,l}\left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l}\pi_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)\left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}^{\ 2}} \right.$$

$$\left. + \frac{n_{(j)a,g,l}\pi_{(j)a,g,l}^{\ 2}\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\ 2}} \right\}$$

$$\frac{\partial^2 \ell}{\partial \theta \partial \xi} =$$

$$\sum_{j=1}^{k}\sum_{l=1}^{3} a \log d_{(j)} \left\{ \frac{-t_{(j)a,g,l}\left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l}\pi_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)\left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}} \right.$$

$$\left. + \frac{n_{(j)a,g,l}\pi_{(j)a,g,l}^{\ 2}\log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}^{\ 2}} \right\}$$

$$\frac{\partial^2 \ell}{\partial \theta \partial \nu} =$$

$$\sum_{j=1}^{k}\sum_{l=1}^{3} g \log d_{(j)} \left\{ \frac{-t_{(j)a,g,l}\left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l}\pi_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)\left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}} \right.$$

$$\left. + \frac{n_{(j)a,g,l}\pi_{(j)a,g,l}{}^2 \log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}{}^2} \right\}$$

$$\frac{\partial^2 \ell}{\partial \theta^2} =$$

$$\sum_{j=1}^{k}\sum_{l=1}^{3} \left(\log d_{(j)}\right)^2 \left\{ \frac{-t_{(j)a,g,l}\left(\log\left(1-\pi_{(j)a,g,l}\right)\right)^2 - t_{(j)a,g,l}\pi_{(j)a,g,l}\log\left(1-\pi_{(j)a,g,l}\right)\left\{1-\log\left(1-\pi_{(j)a,g,l}\right)\right\}}{\pi_{(j)a,g,l}} \right.$$

$$\left. + \frac{n_{(j)a,g,l}\pi_{(j)a,g,l}{}^2 \log\left(1-\pi_{(j)a,g,l}\right)}{\pi_{(j)a,g,l}{}^2} \right\}.$$

Letting:

$$R_{(j)a,g,1} = \frac{n_{(j)a,g,1}\pi_{(j)a,g,1}{}^2 \log\left(1-\pi_{(j)a,g,1}\right) - t_{(j)a,g,1}\left(\log\left(1-\pi_{(j)a,g,1}\right)\right)^2}{\pi_{(j)a,g,1}{}^2}$$

$$- \frac{t_{(j)a,g,1}\pi_{(j)a,g,1}\log\left(1-\pi_{(j)a,g,1}\right)\left\{1-\log\left(1-\pi_{(j)a,g,1}\right)\right\}}{\pi_{(j)a,g,1}{}^2}$$

$$R_{(j)a,g,2} = \frac{n_{(j)a,g,2}\pi_{(j)a,g,2}{}^2 \log\left(1-\pi_{(j)a,g,2}\right) - t_{(j)a,g,2}\left(\log\left(1-\pi_{(j)a,g,2}\right)\right)^2}{\pi_{(j)a,g,2}{}^2}$$

$$- \frac{t_{(j)a,g,2}\pi_{(j)a,g,2}\log\left(1-\pi_{(j)a,g,2}\right)\left\{1-\log\left(1-\pi_{(j)a,g,2}\right)\right\}}{\pi_{(j)a,g,2}{}^2}$$

$$R_{(j)a,g,3} = \frac{n_{(j)a,g,3}\pi_{(j)a,g,3}{}^2 \log\left(1-\pi_{(j)a,g,3}\right) - t_{(j)a,g,3}\left(\log\left(1-\pi_{(j)a,g,3}\right)\right)^2}{\pi_{(j)a,g,3}{}^2}$$

$$- \frac{t_{(j)a,g,3}\pi_{(j)a,g,3}\log\left(1-\pi_{(j)a,g,3}\right)\left\{1-\log\left(1-\pi_{(j)a,g,3}\right)\right\}}{\pi_{(j)a,g,3}{}^2}.$$

The 2nd derivative matrix is then defined as:

$$\begin{pmatrix}
\sum_{j=1}^{k} R_{(j)a,g,1} & 0 & 0 & \sum_{j=1}^{k} R_{(j)a,g,1}\left(\log d_{(j)}\right) & \sum_{j=1}^{k} aR_{(j)a,g,1} & \sum_{j=1}^{k} gR_{(j)a,g,1} \\
0 & \sum_{j=1}^{k} R_{(j)a,g,2} & 0 & \sum_{j=1}^{k} R_{(j)a,g,2}\left(\log d_{(j)}\right) & \sum_{j=1}^{k} aR_{(j)a,g,2} & \sum_{j=1}^{k} gR_{(j)a,g,2} \\
0 & 0 & \sum_{j=1}^{k} R_{(j)a,g,3} & \sum_{j=1}^{k} R_{(j)a,g,3}\left(\log d_{(j)}\right) & \sum_{j=1}^{k} aR_{(j)a,g,3} & \sum_{j=1}^{k} gR_{(j)a,g,3} \\
\sum_{j=1}^{k} R_{(j)a,g,1}\left(\log d_{(j)}\right) & \sum_{j=1}^{k} R_{(j)a,g,2}\left(\log d_{(j)}\right) & \sum_{j=1}^{k} R_{(j)a,g,3}\left(\log d_{(j)}\right) & \sum_{j=1}^{k}\left(R_{(j)a,g,1}R_{(j)a,g,2}R_{(j)a,g,3}\right)\left(\log d_{(j)}\right)^2 & \sum_{j=1}^{k}\left(a\log d_{(j)}\right)\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) & \sum_{j=1}^{k}\left(g\log d_{(j)}\right)\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) \\
\sum_{j=1}^{k} aR_{(j)a,g,1} & \sum_{j=1}^{k} aR_{(j)a,g,2} & \sum_{j=1}^{k} aR_{(j)a,g,3} & \sum_{j=1}^{k}\left(a\log d_{(j)}\right)\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) & \sum_{j=1}^{k} a^2\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) & \sum_{j=1}^{k} ag\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) \\
\sum_{j=1}^{k} gR_{(j)a,g,1} & \sum_{j=1}^{k} gR_{(j)a,g,2} & \sum_{j=1}^{k} gR_{(j)a,g,3} & \sum_{j=1}^{k}\left(g\log d_{(j)}\right)\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) & \sum_{j=1}^{k} ag\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right) & \sum_{j=1}^{k} g^2\left(R_{(j)a,g,1}+R_{(j)a,g,2}+R_{(j)a,g,3}\right)
\end{pmatrix}$$

The expected information matrix is defined as:

$$I_E(\gamma_1,\gamma_2,\gamma_3,\theta,\xi,\nu)=E\left[I_O(\gamma_1,\gamma_2,\gamma_3,\theta,\xi,\nu)\right]=E\left[-\nabla\nabla\ell\right].$$

The asymptotic variance of the function of $\log(TD)$ is:

$$\text{var}\left(\log(TD)+\frac{\xi a+\nu g}{\theta}\right)=$$

$$\nabla\left(\log(TD)+\frac{\xi a+\nu g}{\theta}\right)^{T} I_{E}^{-1}(\gamma_1,\gamma_2,\gamma_3,\theta,\xi,\nu)\nabla\left(\log(TD)+\frac{\xi a+\nu g}{\theta}\right)$$

Where differentiating with respect to $(\gamma_1,\gamma_2,\gamma_3,\theta,\xi,\nu)$ gives the gradient vector of $\nabla\left(\log(TD)+\frac{\xi a+\nu g}{\theta}\right)$:

$$\nabla\left(\log(TD)+\frac{\xi a+\nu g}{\theta}\right)=-\frac{1}{\theta}\left(\frac{e^{\gamma_1}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}},\frac{e^{\gamma_2}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}},\frac{e^{\gamma_3}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}},\log(TD),0,0\right)$$

Pre and post multiplying $I_{E}^{-1}(\gamma_1,\gamma_2,\gamma_3,\theta,\xi,\nu)$ by $\nabla\left(\log(TD)+\frac{\xi a+\nu g}{\theta}\right)$ is equivalent to pre and post multiplying $I_{E}^{-1}(\gamma_1,\gamma_2,\gamma_3,\theta)$ (as in Appendix 3) by $-\frac{1}{\theta}\left(\frac{e^{\gamma_1}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}},\frac{e^{\gamma_2}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}},\frac{e^{\gamma_3}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}},\log(TD)\right)$, which therefore reduces the variance to the same as in Appendix 2, with $\sum_{j=1}^{k} R_{(j)a,g,l}=\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=1}^{1} R_{(j)l}$ in the setting of 2 factors with 2 levels each.

The variance is defined as:

$$\frac{1}{\theta^2 \det}\times$$

$$\left\{\left(\frac{e^{\gamma_1}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)^2\left(\frac{\det+\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)2}\right)\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)3}\right)\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)1}\log d_{(j)}\right)^2}{\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)1}\right)}\right.\right.$$

$$-2\left(\frac{e^{\gamma_1}e^{\gamma_2}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)-2\left(\frac{e^{\gamma_1}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)+2\left(\frac{e^{\gamma_1}\log(TD)}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)$$

$$+\left(\frac{e^{\gamma_2}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)\left(\frac{\det+\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)1}\right)\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)3}\right)\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)2}\log d_{(j)}\right)^2}{\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)2}\right)}\right)$$

$$-2\left(\frac{e^{\gamma_2}e^{\gamma_1}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)-2\left(\frac{e^{\gamma_2}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)+2\left(\frac{e^{\gamma_2}\log(TD)}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)$$

$$+\left(\frac{e^{\gamma_3}}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)\left(\frac{\det+\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)1}\right)\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)2}\right)\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)3}\log d_{(j)}\right)^2}{\left(-\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)3}\right)}\right)$$

$$-2\left(\frac{e^{\gamma_1}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)-2\left(\frac{e^{\gamma_2}e^{\gamma_3}}{\left(e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}\right)^2}\right)+2\left(\frac{e^{\gamma_3}\log(TD)}{e^{\gamma_1}+e^{\gamma_2}+e^{\gamma_3}}\right)$$

$$\left.-\left(\log(TD)\right)^2\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)1}\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)2}\sum_{j=1}^{k}\sum_{a=0}^{1}\sum_{g=0}^{1}R_{(j)3}\right\}$$

11. Bibliography

[1] Postel-Vinay S, Gomez-Roca C, Molife LR, Anghan B, Levy A, Judson I, De Bono J, Soria JC, Kaye S, Paoletti X. Phase I Trials of Molecularly Targeted Agents: Should We Pay More Attention to Late Toxicities? *Journal of Clinical Oncology*, 2011, 29(13):1728-35.

[2] Storer B, Design and analysis of phase I clinical trials. *Biometrics* 1989; 45(3): 925–937

[3] Berry S, Carlin B, Lee J, Muller P, *Bayesian Adaptive Methods for Clinical Trials,* Boca Raton FL Chapman & Hall/CRC Biostatistics Series, 2010

[4] Sinclair K, Whitehead A, A Bayesian approach to dose-finding studies for cancer therapies: Incorporating later cycles of information, *Statistics in Medicine*, 2014; 33: 2665-2680.

[5] Babb J, Rogatko A, Patient specific dosing in a cancer phase I clinical trial, *Statistics in Medicine* 2001; 20:2079-2090.

[6] Skolnik JM, Barrett JS, Jayaraman B, Patel D, Adamson PC. Shortening the Timeline of Pediatric Phase I Trials: The Rolling Six Design, *Journal of Clinical Oncology* 2008;26(2):190-5.

[7] Lee DP, Skolnik JM, Adamson PC: Pediatric phase I trials in oncology: An analysis of study conduct efficiency. *Journal of Clinical Oncology* 2005; 23:8431-8441.

[8] Simon R, Freidlin B, Rubinstein L, Arbuck SG, Collins J, Christian MC. Accelerated Titration Designs for Phase I Clinical Trials in Oncology. *Journal of the National Cancer Institute* 1997; 89(15):1138-1147.

[9] Whitehead J, Brunier H, Bayesian Decision Procedures for Dose Determining Experiments, *Statistics in Medicine* 1995; 14:885-893.

[10] O'Quigley J, Pepe M, Fisher L, Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer, *Biometrics* 1990; 46:33-48.

[11] O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics.* 1996; 52:673–684.

[12] Cheung K, Chappell R, Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities, *Biometrics* 2000; 56:1177-1182.

[13] Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998; 17(10):1103-20.

[14] Whitehead J, Williamson D, Bayesian Decision Procedures Based on Logistic Regression Models for Dose-Finding Studies, *Journal of Biopharmaceutical Statistics* 1998; 8(3):445-467.

[15] Collett D, *Modelling Survival Data in Medical Research*. Second Edition. Boca Raton FL: Chapman and Hall/CRC; 2003.

[16] Zhou Y, Whitehead J, Practical Implementation of Bayesian Dose-Escalation Procedures, *Drug Information Journal* 2003; 37:45–59,

[17] Gibb A, Greystoke A, Ranson M, Linton K, Neeson S, Hampson G, Illidge T, Smith E, Dive C, Pettitt A, Lister A, Johnson P, Radford J, A study to investigate dose escalation of doxorubicin in ABVD chemotherapy for Hodgkin lymphoma incorporating biomarkers of response and toxicity, *British Journal of Cancer* 2013; 109(10):2560-5.

[18] Shen L, O'Quigley J, Consistency of Continual Reassessment Method Under Model Misspecification, *Biometrika* 1996; 83 (2):395-405.

[19] Ferry DR, Smith A, Malkhandi J, Fyfe DW, de-Takats PG, Anderson D, Baker J, Kerr DJ. Phase I clinical trial of the flavonoid quercetin—pharmacokinetics and evidence for in vivo tyrosine kinase inhibition. *Clin Cancer Research* 1996; 2:659–668