Developments to established dose-finding methodologies for application in trials with complex and innovative designs

PhD Proposal

Amit Patel

Supervisors:
Professor Lucinda Billingham
Dr Kristian Brock

Contents

1 Introduction

Early phase clinical trials are essential in the drug development process as they provide key information about new interventions which can be used in later-phase testing. Specifically, phase I trials aim to establish the maximum tolerated dose (MTD) of a new intervention, by determining the dose with the closest dose-limiting toxicity (DLT) probability to the target toxicity level (TTL). Conventionally, designs of these trials are algorithm based, model based or model-assisted based [1]. Some examples of each type of design:

- **Algorithm based designs**

  - 3 + 3
  - Up-and-down

- **Model based designs**

  - Continual reassessment method (CRM)
  - Escalation with overdose control (EWOC)

- **Model-assisted based designs**

  - Bayesian optimal interval (BOIN)
  - Modified toxicity probability interval (mTPI)

Algorithm based designs use pre-determined rules to pick the MTD and allocate patients to a dose level. These are the most common designs implemented in phase I trials due to their simplicity. In contrast, model based designs like the CRM, use a statistical model to determine the dose-toxicity relationship and allocate patients accordingly [2]. Model-assisted designs combine the simplicity of algorithm designs with pre-determined rules for dose escalation and de-escalation with the performance of model based designs, which models data observed at the current dose [1].

Simulations can be used to compare different trial designs by assessing statistical operating characteristics such as: the accuracy of selecting the MTD, percentage of patients assigned to the MTD, percentage of patients over-dosed or under-dosed and the number of DLT's observed. Ananthakrishnan et al. [3] conducted a simulation study assessing a variety of trial designs with comparable design specifications under a range of scenarios. It was found that model based designs performed well and assigned the highest number of patients to the MTD, and also had a relatively high probability of selecting the correct MTD. Some algorithm based designs worked as well as the model based designs but required a higher sample size of patients than those designs. These comparisons were conducted under certain assumptions, and as such different trial designs may be better optimised for different scenarios.

A review into phase I trials of new anticancer therapies found 90 statistical methodology papers and 1235 clinical trials between 1991 and 2006. Only 1.6% (20/1235) of clinical trials used statistical methods presented in the methodology papers, 17 of which were CRM designs and three were EWOC [4]. Another review looking at oncology trials between 2008 and 2014 found 1693 clinical trials, 92 (5.4%) of which used model based/novel designs [5].

The CRM as a design utilises all available trial data and when compared to the 3 + 3, more accurately determines the MTD as well as treating more patients at the MTD. There are a number of key considerations, highlighted by Wheeler et al. [6], that go into this design which require both statistical and clinical input. These are:

- Number of doses

- Target toxicity level

- Dose toxicity model

- Dose toxicity skeleton

- Method of inference

- Sample size and cohort size

- Safety modifications

- Stopping rules / Decision rules

Having too few doses can lead to poor estimation, whilst too many will lead to poor escalation. The target toxicity level is the target probability of a patient experiencing a DLT usually ranging between 20 and 40%. A number of dose toxicity models can be chosen to model the relationship between dose and probability of observing a DLT. The main assumption of this model is that it's monotonically increasing in dose, implying higher doses are more toxic. The dose toxicity skeleton is the DLT probability the investigators expect at each dose level, which is usually determined using clinical input or previous trial data. Two modes of inference can be selected, either a likelihood based approach or a Bayesian approach. A binary toxicity outcome is observed from each patient, they either have a DLT or they do not. This data is then used in the dose toxicity model along with the dose toxicity skeleton to produce a posterior distribution which can be used to infer probability of DLT at each dose. The dose closest to the TTL becomes the recommend dose for the next patient/cohort. Sample size is usually constrained by practical issues such as the number of centres involved, disease prevalence and recruitment rates. Simulations can be conducted in order to optimise operating characteristics under feasible sample sizes. Using cohorts of one patient would theoretically provide the best results however, this may potentially be harmful to patients due to quick escalation without observation of necessary safety data. Safety modifications can be introduced to further reduce the chance of overdosing patients. Two common examples are starting the trial at a lower dose and not escalating past untried doses. A number of stopping rules can be implemented to reduce trial duration before the maximum sample size is reached. Criteria should be established for scenarios where: the MTD appears to be outside the range of pre-defined doses, the lowest dose is too toxic, sufficient evidence the MTD has been reached early or a combination of these [6].

The CRM design has been extended and modified to deal with a number of issues that arise in phase I trials. Certain interventions, such as radiotherapy or some immunotherapy, may cause late-onset toxicities. Normal CRM designs would have to increase the duration of the DLT evaluation period to capture these toxicities, which would inevitably prolong the trial. The time-to-event CRM, referred to as the TITE-CRM developed by Cheung and Chappel [7], presents a solution by considering partial information on patients. If a patient is part way through their evaluation without experiencing a DLT, they are evaluated as having tolerated the dose up to the observed time-point. This is incorporated into the CRM model with a weighting parameter calculated for each patient as a function of time. Patients who reach the end of the DLT evaluation period with no toxic reaction are considered to have tolerated the dose and contribute full information to the model. Similarly, those who experience a DLT at any time also contribute fully to the model. The partial information only applies to tolerances that have been observed part way through the evaluation period. The TITE-CRM allows for patients to be entered into the study in a staggered fashion without compromising accuracy of estimation for the MTD, whilst also significantly reducing trial duration. However, in scenarios where recruitment is rapid the TITE-CRM could allocate a number of patients to a dose level which

hasn't been observed for a significant period of time. This could lead to reduced efficiency in the scenario where the dose level is truly safe as higher doses aren't being explored. Many patients could also be exposed to a higher toxicity dose if the current dose exhibits multiple late-onset toxicities. Both of these issues can be resolved by allocating dose levels to a cohort of patients and waiting till the DLT evaluation period has passed to allocate the subsequent dose level [8].

For immunotherapies and certain molecularly targeted agents it may be inadequate to consider only toxicity and tolerability in dose finding trials. Scenarios may occur where higher doses may not be more efficacious than lower doses, DLT's may not be observed before escalation and toxicity may not be a surrogate for activity. In these cases it would be important to consider efficacy as an additional end-point. Other designs such as the bivariate CRM (bCRM) and the EffTox design model bivariate binary and trinary outcomes respectively, which aims to recommend a dose level not only based on a drugs toxicity but its efficacy as well. The bCRM, in comparison to the traditional CRM, defines the MTD as a function of two outcomes toxicity and response which are combined into a joint likelihood model, an additional parameter is also included to account for the within-subject association of both outcomes [9]. Alternatively, EffTox a Bayesian adaptive phase I / II design by Thall and Cook [10], considers the trade-offs between toxicity and efficacy. This method allows for binary as well as trinary outcomes where efficacy and toxicity are disjoint and neither can occur. The optimal dose is selected based upon the non-Euclidean distance between the ideal point (that being (1, 0) in the (prob-eff, prob-tox) quadrant which is the point where efficacy is guaranteed and toxicity impossible) and the expected probabilities of efficacy and toxicity given the data, based on a family of contours characterising these trade-offs for each dose [10]. EffTox designs are best suited when there is an early biomarker for efficacy available and the disease population for the drug is homogeneous.

The general objective of early-phase trials is to identify/recommend promising dose regimens for investigations in phase III trials. Phase I trials are dose finding that examine safety and toxicity and evaluate an MTD for use in phase II trials which aim to find evidence of efficacy. Independently these trials can take an incredible amount of resources to set-up and conduct. Adaptive designs such as EffTox can be used to facilitate seamless transitions between phase I / II trials and expedite the process. Furthermore, the use of early-phase platform trials could be used to facilitate such designs. Platform trials are typically randomised trials in a single histology which investigates multiple biomarkers and multiple treatments [8]. Rather than answer a specific clinical question the trial aims to randomise treatments to biomarker strata. Designs such as EffTox and issues such as delayed toxicities should always be considered when designing early-phase trials. Platform trials effectively take advantage of designs such as these as within -trial information of efficacy and safety can be shared.

2   Aims

The main aims of this project are to investigate extensions to existing methodologies in early-phase trials with complex and innovative designs. This will involve introducing new concepts for use in trials, adapting current methodology for use in platform trials and extending certain early-phase designs. Another aim will be to explore the application and implementation of these extensions, into trials currently being conducted, in order to gauge the effectiveness of these developments.

3 Topics and Chapters

3.1 Applications and implementation into real world trials

Most of the proposed topics revolve around methodological developments which can be made to improve and enhance the conduct of early phase trial designs. However, it is also important to consider how these developments can be implemented effectively and their impact on actual trials. There are two trials, which are yet to be open, in which some of this work could be put to use and evaluated.

Radiant-BC

Radiant-BC is a multi-arm phase Ib/expansion cohort platform trial of the efficacy and safety of radiosurgery with durvalumab (immunotherapy) and systemic therapies as part of standard of care in patients with brain metastases secondary to breast cancer.

This trial involves three groups each allowing for stratification into trial arms to the clinicians' choice of systematic therapies. The platform trial design allows for the addition of extra arms into the groups. Heterogeneity within the arms of the groups may become an issue. Statistical methodology to deal with these issues is a topic of interest.

ADePT-DDR

ADePT-DDR is an open-label, multi-centre, platform trial which will evaluate the safety and efficacy of different DNA Damage Repair (DDR) agents, or different immunotherapy agents and/or DDR and immunotherapy combinations, together with radiotherapy in patients with head and neck squamous cell carcinoma being treated curatively. The trial will allow for the evaluation of new DDR agents and for roll-on to a randomised phase II / III trial. The initial component of this trial will open with a single-arm dose-finding phase Ib / IIa trial.

The initial component of this trial is a dose escalation design based on a two-stage modified TITE-CRM with partial ordering. As a combination of interventions are being investigated we have a scenario where dose toxicity is not necessarily monotonically increasing, which is a fundamental assumption of the dose toxicity model. Methodology has been introduced to deal with this problem [11] however, this may be the first time it has been implemented into a trial. As such the conduct of this trial may be of interest.

3.2 Literature review

The most recent literature review pertaining to methodologies used in early-phase trials was published in 2014. Since then new methodologies have emerged and new tools to make conducting these trials more accessible. Updating this knowledge will be useful in understanding the current number of early-phase trials utilising statistical designs and help identify if previous challenges have been addressed or if any new barriers to entry have developed. Early-phase platform trials could also be included in this review to see how many examples exist.

This information could then be used to establish comprehensive guidance of available methodologies and optimal scenarios in which to use them. Simulations can be incorporated to compare operating characteristics for various scenarios to validate the guidance provided.

3.3 Entropy classification

Entropy, also referred to as Shanon entropy, is a term from information theory and is defined as a measure of unpredictability of information content. It is essentially a measure of uncertainty. The entropy formula for an event X with n possible outcomes and probabilities $(p_1, ..., p_n)$ is

$H(X) = H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$. Entropy is maximum when all outcomes are equally likely. Any time we move away from equally likely outcomes or introduce predictability entropy goes down.

In the context of dose-finding trials we have a number of doses which could be classified as the MTD. Entropy as a tool may be useful as a measure of uncertainty. The application of this could be applied retrospectively to a trial, or to a simulated trial, to investigate how entropy changes over its duration. As we collect more information we would expect uncertainty to decrease implying the entropy would decrease. This could be used to establish thresholds for terminating trials early if a certain level of entropy is reached. In simulations entropy could be utilised as an operating characteristic to compare uncertainty under different design specifications. It could also be used to help inform simulations for sample size decisions by calculating the change in uncertainty additional patients are likely to bring.

3.4   Issues of heterogeneity

Designs also exist for scenarios where there is expected patient heterogeneity, where patients can be put into groups and assessed in the same study rather than running separate trials in parallel which may be infeasible. The simplest case is when patients can be separated into two separate groups. Methodology exists to deal with this (CRM for ordered groups [12], two-sample CRM [13]). Real scenarios can obviously be more complex, there may be multiple groups or continuous covariates to consider and due to sample size limitations different tools are needed to deal with this. More recently designs for partially ordered groups have been published. This is for scenarios where only some of the groups can be ordered with respect to the probability of toxicity at that dose. One method by Conaway [14] makes use of parametric models in combination with the Hwang-Peddada order-restricted estimation procedure. Another, proposed by Horton et al. [15], uses shift models. Hierarchical modelling has been suggested by Cunanan and Koopmeiners [16,17] as a way to evaluate the MTD in multiple patient populations, simulations have been conducted in both CRM [16] and efficacy/toxicity [17] designs and show improvements in operating characteristics. In the context of early-phase platform trials this will be an area to explore as information can be shared across heterogeneous population groups. Implementation of these designs/methods may be complex as such it would be beneficial to see real world scenarios where these are successfully applied. Alternative methodologies or software may be developed to help improve issues of heterogeneity.

3.5   Applications and packages

Some methods and designs require a lot of input and are difficult to implement. Web applications allow both statisticians and clinicians a better understanding of how certain designs work and make them more accessible. These applications are generally limited in terms of making modifications to the design the app features. Packages are more robust and provide access to code in order to make modifications for complex trial designs. Due to the number of designs and modifications to those designs there are a large amount of packages available however, it makes trying to combine all of these things into one place difficult. Most web applications seem to be based on the CRM design so applications for more niche designs could be feasible.

3.6   Extension to the Wages and Tait design

Wages and Tait [18] developed an adaptive phase I / II trial design which considers information on both toxicity and efficacy. The design aims to find the optimal biological dose (OBD) and assess toxicity in a similar way to the CRM to determine doses with DLTs probabilities under the TTL. Using these subset of doses the model then selects the most efficacious of the doses,

initially done with random allocation then by using estimates from the model.

This design could be further extended by the inclusion of a zero dose/control dose to create a control group in order to compare efficacy outcomes. To begin the OBD is selected randomly from a subset of doses the model deems safe. At this point it should be possible to be allocated to the control dose. Rules should be implemented to determine the probability of the control dose being selected however this would require further investigation to determine a logical answer. This will require implementation, coding and estimation of performance. This design can also be thought of in the context of an early-phase platform trial and how it could be implemented.

3.7 Other areas of research

1. Implementation of partial ordering into the ADePT-DDR trial

   - Partial ordering methodology for scenarios where dose toxicity isn't monotonically increasing has been developed and extended to include a TITE component. However, this is perhaps the first instance of its implementation as such lessons learnt from this trial may be of value.

   - Code had to be developed in order to conduct simulations and the trial itself. This may be of value to others hoping to conduct a trial using similar methodology.

2. Optimising model specification/dose escalation/sample size

   - Simulations are currently used as the main method of determining sample size. Model specifications are often based on clinical input, however it seems like most of this is guess work. Extensions to simulations could be investigated to see if these could be improved.

   - Some methods exist for calculating sample size but doesn't seem like they are widely implemented. These methods are based on CRM designs and it would be useful to extend them for other designs and consider their implementation into a real trial.

3. Modifying trial design whilst its still open (inserting/removing dose levels)

   - Initial guesses of the dose toxicity curve may be incorrect. Doses may be much closer/further in terms of toxicity which would motivate the removal/addition of dose levels. Could investigate the use of pre-specified criteria to establish rules of when new dose levels should be added/removed. Possibly use the dose toxicity curve to predict what the optimal dose level would be based on the target DLT rate.

   - Simulations of trial designs which can alter dose levels would become more complex.

   - This could also be looked at for model-assisted based designs which model data around the current dose level.

   - The feasibility of this could also be investigated in seamless phase I/ II designs such as EffTox and Wages and Tait's design.

7

4   References

References

[1] Heng Zhou, Thomas A. Murray, Haitao Pan, and Ying Yuan. Comparative review of novel model-assisted designs for phase I clinical trials. *Statistics in Medicine*, 37(14):2208–2222, 2018.

[2] Sharon B. Love, Sarah Brown, Christopher J. Weir, Chris Harbron, Christina Yap, Birgit Gaschler-Markefski, James Matcham, Louise Caffrey, Christopher McKevitt, Sally Clive, Charlie Craddock, James Spicer, and Victoria Cornelius. Embracing model-based designs for dose-finding trials. *British Journal of Cancer*, 117(3):332–339, 2017.

[3] Revathi Ananthakrishnan, Stephanie Green, Mark Chang, Gheorghe Doros, Joseph Massaro, and Michael LaValley. Systematic comparison of the statistical operating characteristics of various Phase I oncology designs. *Contemporary Clinical Trials Communications*, 5:34–48, 2017.

[4] André Rogatko, David Schoeneck, William Jonas, Mourad Tighiouart, Fadlo R. Khuri, and Alan Porter. Translation of innovative designs into phase I trials. *Journal of Clinical Oncology*, 25(31):4982–4986, 2007.

[5] Cody Chiuzan, Jonathan Shtaynberger, Gulam A. Manji, Jimmy K. Duong, Gary K. Schwartz, Anastasia Ivanova, and Shing M. Lee. Dose-finding designs for trials of molecularly targeted agents and immunotherapies. *Journal of Biopharmaceutical Statistics*, 27(3):477–494, 2017.

[6] Graham M. Wheeler, Adrian P. Mander, Alun Bedding, Kristian Brock, Victoria Cornelius, Andrew P. Grieve, Thomas Jaki, Sharon B. Love, Lang'O Odondi, Christopher J. Weir, Christina Yap, and Simon J. Bond. How to design a dose-finding study using the continual reassessment method. *BMC Medical Research Methodology*, 19(1):1–15, 2019.

[7] Ying Kuen Cheung and Rick Chappell. Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities. *Biometrics*, 56(4):1177–1182, 2000.

[8] Mei-Yin C. Polley and Ying Kuen Cheung. Early-Phase Platform Trials: A New Paradigm for Dose Finding and Treatment Screening in the Era of Precision Oncology. *JCO Precision Oncology*, (3):1–8, 2019.

[9] Thomas M. Braun. The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials*, 2002.

[10] Peter F Thall and John D Cook. Dose-Finding Based on Efficacy – Toxicity Trade-Offs. *Biometrics*, (September):684–693, 2004.

[11] Nolan A. Wages, Mark R. Conaway, and John O'Quigley. Using the time-to-event continual reassessment method in the presence of partial orders. *Statistics in Medicine*, 32(1):131–141, 2013.

[12] John O'Quigley and Xavier Paoletti. Continual reassessment method for ordered groups. *Biometrics*, 59(2):430–440, 2003.

[13] J. O'Quigley, L. Z. Shen, and A. Gamst. Two-sample continual reassessment method. *Journal of Biopharmaceutical Statistics*, 9(1):17–44, 1999.

[14] Mark R Conaway. A design for phase I trials in completely or partially ordered groups. 36(15):2323–2332, 2018.

[15] Bethany Jablonski Horton, Nolan A. Wages, and Mark R. Conaway. Shift models for dose-finding in partially ordered groups. *Clinical Trials*, 16(1):32–40, 2019.

[16] Kristen M. Cunanan and Joseph S. Koopmeiners. Hierarchical models for sharing information across populations in phase I dose-escalation studies. *Statistical Methods in Medical Research*, 27(11):3447–3459, 2018.

[17] Kristen M. Cunanan and Joseph S. Koopmeiners. Efficacy/toxicity dose-finding using hierarchical modeling for multiple populations. *Contemporary Clinical Trials*, 71(June):162–172, 2018.

[18] Nolan A. Wages and Christopher Tait. Seamless Phase I/II Adaptive Design for Oncology Trials of Molecularly Targeted Agents. *Journal of Biopharmaceutical Statistics*, 25(5):903–920, 2015.