Final Project: Bayesian Inference Clinical Trials and Nonparametric Models

Non-stat Team: Captain: Xinyi Zhang, Members: ...

2023-05-21

1. Introduction

- Background of the study

2. Literature Review

2.1 CRM design for phase 1 dose-finding trials

CRM is an alternative to the standard 3+3 design based on using a model like a one parameter or two parameter logistic model, to understand the maximum tolerated dose in a phase 1 trial. CRM is more accurate in choosing the MTD, is less likely to choose ineffective doses, treats fewer patients at overly toxic doses, and treats fewer patients at very low doses.(Garrett-Mayer 2006) Our paper plans to look at a two-parameter model. A two-parameter model is likely to better estimate the shape of the entire dose-toxicity relationship, [34] but less efficiently identify the MTD; it may take longer to reach the MTD since two parameters must be estimated, and there may be difficulties fitting the model or obtaining consistent estimates of model parameters. (Iasonos et al. 2016)

The idea behind the CRM starts with a priori dose toxicity curve and a chosen target toxicity rate. This curve will be refit after every cohort (1-3 patients) toxicity outcome, is observed. At every new dose or same dose, the all-prior data is used to update the model/curve (Garrett-Mayer E). As required a discussion needs to take place with all relevant stakeholders. The target toxicity level is typically set between 20 to 25% and can be as high as 40%. (Brock et al. 2017; Møller 1995) In a review of 197 phase I trials published between 1997 and 2008, the median number of dose levels explored was five (range 2–12). (Penel and Kramar 2012)

Inference or decisions can be made using a likelihood or Bayesian methods using the accruing trial data and clinical judgment. In a Bayesian method data from patients in the trial is used to update prior on the model distribution which then gives a posterior distribution for the model parameters and therefore posterior beliefs for the probability of DLT at each dose. These posterior probabilities are used to make dose escalation decisions. By assessing a design's operating characteristics with a specific prior in a variety of scenarios, the prior distribution can be recalibrated until the model makes recommendations for dose escalations and the MTD that the trial team are happy with (Wheeler et al. 2019).

Possible decision rules include choosing the dose with an estimated probability of DLT closest to the TTL or, more conservatively, choosing the dose with an estimated probability of DLT closest to, but not greater than, the TTL. The first option allows quicker escalation towards the true MTD but may expose more patients to overdoses. The second option reduces the chance of overdosing patients, but may take longer to escalate towards the true MTD (Wheeler et al. 2019)

Samples sizes are determined by the study and how and where its being conducted. Specifying a lower bound based on Cheung's work and practical upper bound in trial protocols. Cheung [45] proposed formulae that

use a target average percentage of correctly selecting the MTD (say, 50% of the time) to obtain a lower bound for the trial sample size (Wheeler,M Graham). Although CRM designs, like standard ones, can halt after only 10–14 subjects, it is typically necessary to plan for at least 18–24 total subjects, after which the probability of a correct MTD choice rises slowly with sample size (Garrett-Mayer 2006)

Cohort size at each dose level typically is more than 1. A cohort size of one allows better understanding of operating characteristics but this is rarely used. (Garrett-Mayer 2006) There could be regulatory constraints. If cohort size is greater than 2, then a monitoring plan is needed.

Stopping rules for the trial include the following examples. Early termination can be considered if the MTD is judged to be outside the planned set of doses. Adding additional patients is unlikely to yield information that would change the current MTD estimate. Fixed no of patients has been consecutively dosed at one dose level. Estimated probability of all dose levels having a DLT rate about the TTL is at least 90%. The probability that the next m patients to be dose in the trial will be given the same dose levels, regardless of DLT outcomes observed, exceed some level (Wheeler et al. 2019).

2.2 Summary for references

2.2.1 References 1-5 (by Tongtong Jin)

Phase 1 trial and maximum tolerated dose (MTD) The phase 1 trials in oncology are usually designed to obtain the optimal dose of a new treatment for efficacy testing in subsequent phase 2 trials. For cytotoxic agents, the probability of treatment benefit is presumed to be positively proportional to the dose in a certain range of consideration. Thus, the optimal dose in phase 1 trial is usually considered as the highest dose at a tolerable level of toxicity. And the optimal dose we are seeking for is exactly the maximum tolerated dose (MTD).

To define MTD in a more rigorous way, it is the dose expected to produce some degree of medically unacceptable, dose limiting toxicity (DLT) in a specified proportion θ of patients. Namely,

$$Prob(DLT|Dose = MTD) = \theta,$$

where the proportion θ is also defined as the target toxicity level (TTL) (Babb and Rogatko 2004; Cater 1972).

Dose escalation methods in phase 1 trials To find the MTD we defined above, in clinical testing we adopt dose escalation methods, which is based on the prior belief that the toxicity increases monotonically with increasing dose. The principle of dose escalation in phase 1 trials is both maintaining the toxicity at a safe level and the information accumulation at a rapid speed and at the same time avoid patients being exposed to subtherapeutic doses as much as possible.

Dose escalation methods can be mainly classified in two branches, rule-based designs such as traditional 3+3 design and model-based designs such as continual reassessment method. Rule-based designs don't make any assumptions for the function of toxicity with respect to dose level. And the next step of dose is purely dependent on the information from the last dose. Then finally terminates at some certain stopping criteria. But model-based designs assume there's a specific function between dose and toxicity, usually power functions, logistic functions etc., and then apply accumulated information from every dose to determine the next dose.

From the perspective of practical use, rule-based designs like traditional 3+3 are easier to implement, but model-based designs need biostatistical expertise and available software on site to perform real-time model fitting. As for the information utilization, rule-based designs only use current information, but model-based designs make use of all toxicity information accumulated during the trial, which can be more comprehensive. In the aspect of the exposure to subtherapeutic doses, model-based designs relatively treat fewer patients at suboptimal doses than rule-based ones. Hence according to the principle of dose escalation methods, the

model-based designs usually do better in rapid information accumulation and reducing excessive exposure to subtherapeutic doses.(Le Tourneau, Lee, and Siu 2009).

Current popularity of rule-based designs and model-based designs in phase 1 trials Although the model-based designs show great advantages in many aspects, rule-based designs like the 3+3 design are still more prevailingly used, and model-based designs are rarely used. Some statistical results about the popularity of these two types of methods in phase 1 trials are as follows.

Rogatko et al. (2007) examined through the records of cancer phase 1 trial from the Science Citation Index database from 1991 to 2006 and divided them into two sets (dose-finding trials and methodologic studies of dose-escalation designs). Then track among these two sets which trials adopted new statistical designs. As a result, only 1.6% trials follow one of the methodologic studies and show extensive lags on publication time. The rest of the trials all follow the traditional up-and-down method (a type of rule-based method).

Chiuzan et al. (2017) studied the degree of adoption of the methods with new trial designs on early phase trials of molecularly targeted agents (MTA) and immunotherapies. It searched papers published from 2008 to 2014 about phase 1 oncology trials and found that in dose-finding trials, 92.9% of them utilized rule-based designs and 5.4% used model-based designs or other novel designs. Particularly, among the MTA and immunotherapies trials, 5.8% used model-based designs. The results show that the adoption of model-based designs and novel designs remains low.

Above phenomenon could be caused by limited time and effort of clinicians and statisticians and the lack of comprehensive and detailed tutorials and instructions for the newly designed approaches.(Rogatko et al. 2007; Chiuzan et al. 2017).

2.2.2 References 6-10 (by Yiwei Ding)

Reference 6: O'Quigley and Zohar (2006)

For the atistical design of dose-finding studies, the standard design is a 'memoryless' design and it's not so satisfying. This paper describes designs with memory and we discuss how these designs are superior to memoryless designs. The most well-known design with memory is the continual reassessment method (CRM).

Reference 7: Harrington et al. (2013)

Carrying out dual-agent phase I trials for medications is crucial. There are predominantly two kinds of dose-escalation trials: rule-based and model-based. Trials based on models are progressively adjusted with the help of Bayesian techniques, which merge preliminary data concerning the dose-toxicity relationship. Studies using simulations indicate that model-driven designs tend to treat a greater proportion of patients at near-optimal dose levels.

Reference 8: A quick guide why not to use A+B designs

- Note: paper to be added

This paper summarize why modelbased designs such as the continual reassessment method (CRM) are more available than 3+3 and similar rulebased A+B designs. Compared with rule-based designs, Modelbased designs can clearly define and can flexibly choose target DLT rate; many patients can be treated at the optimal dose; few patients can be treated at subtherapeutic doses; the utilization of available data is efficient; extension to more complex questions is smooth & straightforward; deviations from the plan are easily accommodated.

Reference 9: Jaki, Clive, and Weir (2013)

There are three classes of dose-escalation trial design: gorithmic approaches (including the popular 3+3 design), Bayesian model-based designs and Bayesian curve-free methods. The main benefit of algorithmic approaches is the simplicity. Model-based and curve-free Bayesian approaches are more preferable because they are more able to identify the dose with the desired toxicity rate and allocate a greater proportion of patient. For statistical and practical reasons, Bayesian model-based or curve-free approach is better. If there is sufficient evidence of high enough quality from previous studies, the model-based approach will be better, otherwise curve-free one is better.

Reference 10: O'Quigley, Pepe, and Fisher (1990)

For the design and analysis of Phase I clinical trails in cancer, attention focuses rather on identifying a dose with a given targeted level is the best estimate of this level. Such sequential designs is called continual reassessment method (CRM). In the procedure, we update our notion of the dose-reponse relationship. From the simulations, this method is good.

2.2.3 References 11-15 (by Ujjwal Sehrawat)

Clinicians are very interested in phase 1 dose finding designs that can estimate the MTD (Maximum Tolerated Dose) using fewer patients with a fixed number of doses, or can test more dose levels for a given sample size. Many studies have compared the CRM (Continual Reassessment Method) and its variants to the traditional SM (3+3 standard method that escalates doses after 3 patients with an option for an additional 3 patients; also called traditional method TM or the "up-and-down scheme proposed by Storer,1989) and found that CRM is more likely to recommend the correct MTD and dose more trial patients close to the MTD. (O'Quigley 1999; Thall and Lee 2003; Iasonos et al. 2008; Onar, Kocak, and Boyett 2009; Onar-Thomas and Xiong 2010)

O'Quigley et al. (the creator of the original CRM) in his present paper (O'Quigley 1999) refuted Korn et al's conclusions about CRM vs SM in terms of the former's purported relatively poor time duration and lack of safety. They did this by rerunning Korn et al's analysis (a set of simulations for 3 different dose toxicity situations) with a few important adjustments. According to O'Quigley, Korn et al. didn't use the actual CRM but a modified version by incorporating an ad hoc stopping rule which disqualifies their modified version of CRM to belong to CRM class of models as it fails to satisfy elementary conditions of CRM's original specification (given by O'Quigley) which recent work by Shen and O'Quigley also showed must be respected in order to demonstrate convergence to the correct dose level. Also, Korn et al. is criticizes for using significantly different sample sizes when comparing CRM and SM which was characterized as problematic and unnecessary. O'Quigley in his reanalysis, instead, used a fixed sample scheme, taking sample size for CRM to be the nearest integer to the average sample size of SM scheme and used the original model presented by O'Quigley et al, thus reaching completely opposite conclusions from Korn et al. Trial Duration: Korn et al. did not consider the 3 by 3 grouped inclusions for its CRM, hence making it incomparable with the 3+3 SM design. Trial duration can, therefore, be seen to be simply a function of design and indeed CRM does not take 2.5 longer time than SM in a fair balanced comparison. Safety: Korn et al. said that CRM treats 70% more patients at dose levels 2 higher than the MTD than the standard method. O'Quigley argued that the "70%" number was an indirect calculation for which the higher dose levels in question used 12% by modified CRM and 7% by standard method. This can be seen as not a great difference and a misleading interpretation by Korn et al. It is, indeed, safer for a randomly chosen patient to be included in a CRM design than a standard 3+3 design. The probability of being treated at very high toxic levels is almost always higher with the standard design than the CRM design during simulations. CRM also performs much better than SM if treating patients at unacceptably low sub-therapeutic levels is considered part of the safety definition. Unlike 3+3, it is entirely straightforward to adjust CRM to make it safe as we require. All it requires is to change the target level, say from 0.2 to 0.1. In this case, the observed number of toxicities will be, on average, roughly halved. Flexibility & Future Scope: One of the main advantages of CRM is its flexibility and ability to be adapted to potentially different situations unlike the standard model which is rigid, samples independently

of any targeted percentile, and has no convergence properties. CRM targets some given toxicity level and concentrates experimentation around that level, using only the information available for independent toxic responses (yes/no) given the dose level. Subsequent progress in the area is likely to come from techniques which incorporate additional information such as pharmacokinetics, group heterogeneity, graded toxicities and intra-patient studies. None of these potential developments can be addressed in a constructive way via the standard method. It is also important to note that the paper's comparison parameters are driven by the standard method, parameters that have often placed the method in an artificially favorable light. (O'Quigley 1999)

Thall et al describes and compares 2 practical outcome-adaptive statistical methods for dose finding in phase 1 clinical trials: CRM and a logistic regression model-based method. Both methods use Bayesian probability models as a basis for learning from the accruing data during the trial, choosing doses for successive patient cohorts, and selecting an MTD. These methods are illustrated and compared to the SM by application to a particular trial in renal cell carcinoma. The paper compares their average behavior by computer simulation under each of several hypothetical dose-toxicity curves. The comparisons show that the Bayesian methods are much more reliable than the conventional algorithm for selecting an MTD, and that they have a low risk of treating patients at unacceptably toxic doses. (Onar, Kocak, and Boyett 2009)

Iasonos et al. compares several CRM-based methods with SM in terms of the number of patients needed to reach the MTD, total sample size required, and trial duration under different scenarios using 2 alternative schemes: a fixed vs a varying sample approach with the implementation of a stopping rule (stopping rule halts the trial if the Confidence Interval around the MTD is within a pre-specified bound). For comparison, variations were given to the number of dose levels (5 to 8) and the location of the true MTD. Only CRM with constraint in dose escalation was evaluated since it is more likely to be used by clinicians as the O'Quigley's original CRM allows skipping dose levels in the absence of DLT's, potentially unnecessarily exposing patients to highly toxic drug levels which makes clinicians uncomfortable. Furthermore 3 CRM-based methods that combine rule-based and model-based approaches were evaluated. Results are generalizable to only CRM designs of this type and not others. They found that CRM and SM are comparable in terms of how fast they reach the MTD as well as the total sample size needed when testing a limited number of dose levels (<=5), however, as the number of dose levels was increased, CRM reached the MTD in fewer patients when used with a fixed sample of 20 patients. However, a sample size of 20—25 patients is not sufficient to achieve a narrow precision around the estimated toxicity rate at the MTD. CRM with a fixed-sample performed better than a CRM with stopping rule that ensures a narrow confidence interval around the toxicity rate at the MTD. CRM-based methods were found to be better than SM in terms of accuracy and optimal dose allocation in almost all cases except when the true dose is among the lower levels. (Iasonos et al. 2008)

Onar et al. provides modifications to CRM largely motivated by specific challenges encountered in context of the Pediatric Brain Tumor Consortium trials, study operating characteristics of this modified CRM through simulations, and compare it to SM. Modifications to the CRM were as follows: Some versions of CRM assume availability of doses in a continuous way given a range, this paper uses preset levels as it is more acceptable to clinicians, easier to manage operationally especially in multi-institutional settings. Patients in these pediatric trials were dosed by body surface area (BSA) instead of in terms of "mg" as done in adult trials. A frequentist likelihood-based approach with a 2-parameter logistic model (Piantadosi et al,1998). Phu(x_j, a) = exp(alpha + beta* x_j) / (1 + exp(alpha + beta* x_j}). The 2 parameter logistic model was favored over its 1 parameter counterpart due to its flexibility even though more information is needed to identify the parameters. Also, "prior information" (using prior suggested by Piantadosi, 1998) was used to fit the model which is needed especially during early trial stages. Compared to SM, simulations indicate that their modified CRM was more accurate, exposed fewer patients to potentially toxic doses and tended to require fewer patients. They also cite the fact that the CRM-based MTD has a consistent definition across trials as one of its many advantages as it is important especially in consortium settings where multiple agents are being tested in studies often running simultaneously and accruing form the same patient population. Note that these results are only generalizable to pediatric trials with BSA adjusted dosing. (Onar, Kocak, and Boyett 2009)

Onar-Thomas et al. compared the performance of CRM vs SM vs the Rolling-6 designs via simulations with respect to overall toxicity, sample size, % of patients treated at doses above the MTD (safety), trial duration

and the dose chosen as the MTD. The Rolling-6 design is a relative newcomer developed with the intention to shorten trial duration by minimizing the period of time during which the trial is closed to accrual for toxicity assessment.

Results indicate that the toxicity rates are comparable across the 3 designs, but the SM and the Rolling 6 designs tend to treat a higher % of patients at doses below the MTD. In cases where 5,6 or more dose levels are proposed to be studied and some toxicities are expected, model-based designs (CRM) have distinct advantages in being able to use the data from all dose levels in estimating the MTD, in accommodating patient-specific dosing and in providing an MTD estimate that is associated with a toxicity probability. Doses identified as MTD by the SM and the Rolling-6 differ in a large % of trials. Results also show that body surface area (BSA) - based dosing used in pediatric trials can make a difference in dose escalation/de-escalation patterns in CRM relative to cases where such variations are not taken into account in the calculations (such as most adult trials) also leading to different MTDs in some cases. Simulation-based comparisons showed that Rolling-6 lead to shorter trials relative to SM. CRM lead to shorter trial duration for slow to medium accrual rates whereas Rolling-6 may have an advantage if the accrual rate is fast. Rolling-6 may be preferable over the CRM if very few or no toxicity is expected with the agent under study and if the dose finding period is long. (Onar-Thomas and Xiong 2010)

2.2.4 References 16-20 (by William)

Reference 16: Le Tourneau et al. (2012)

This paper provide evidence that more extensive implementation of innovative dose escalation designs such as mCRM and ATD in phase I cancer clinical trials of molecularly targeted agents. This paper did a literature review based on 84 trials that reached MTD. The goal was to get more insight on the efficiency of new dose escalation methods in phase I trials of molecularly targeted agents. The literature review indicated a standard 3+3 design was used in 41 trials (49%) while newer algorithm based methods were also used, including ATD in 35 trials (42%) and CRM (mCRM), which was employed in only 6 trials (7%). The mean MTD to starting dose ratio appeared to be at least twice as high for trials using a mCRM or an ATD as for trials using a standard "3+3" design. The mean number of patients exposed to a dose below the MTD for all three trial designs was similar, ranging from 19 to 23. the mean number of patients exposed to doses exceeding the MTD was at least twice as high in trials using a standard "3+3" design or an ATD when compared to trials using a mCRM

Reference 17: Garrett-Mayer (2006)

The paper 'The continual reassessment method for dose-finding studies: a tutorial' provides detailed explanation how to implement a CRM. Does simulation with varying parameters to show how the CRM is better than the 3+3

Reference 18 (Preview version): Cheung (2011)

Reference 19 (Preview version): O'Quigley, Iasonos, and Bornkamp (2017)

Reference 20: Food, Administration, et al. (2013)

FDA guidance indicates Bayesian methods can be used in clinical trial for Drugs and Biologics. This guidance indicates that Bayesian inference is characterized by drawing conclusions based directly on posterior probabilities that a drug is effective and has important differences from frequentist inference (Berger and Wolpert 1988). For trials that use Bayesian inference with informative prior distributions, such as trials that explicitly borrow external information, Bayesian statistical properties are more informative than Type I error probability.

3. Methodology

3.1 Describe the use of the logistic regression model in our study (By Tongtong Jin)

The model we use here to predict the relation between dose level and the probability of DLT is the two-parameter logistic regression model which has the form as follows:

$$p_j = p(d_j|\beta_1, \beta_2) = \frac{exp(\beta_1 + exp(\beta_2)d_j)}{1 + exp(\beta_1 + exp(\beta_2)d_j)}.$$

And in Bayesian setting, this is the likelihood function for dose level j given β_1 and β_2.

3.1.1 Bayesian CRM

In the Bayesian setting of CRM, we first need to choose the prior distributions of parameters β_1 and β_2. Let's denote the prior by $f(\beta_1, \beta_2)$. Then the posterior distribution given the data of k dose levels D_k is as follows:

$$L(D_k|\beta_1, \beta_2) = \prod_{j=1}^{k} p_j^{y_j}(1 - p_j)^{n_j - y_j},$$

where n_j is the number of tested patients at the j-th dose level, and y_j is the number of patients showing DLT at the j-th dose level. Then we can get to the posterior distribution given D_k by applying above in the Bayes' rule. The posterior is

$$p_k(\beta_1, \beta_2|D_k) = \frac{L(D_k|\beta_1, \beta_2)f(\beta_1, \beta_2)}{\iint L(D_k|\beta_1, \beta_2)f(\beta_1, \beta_2)d\beta_1 d\beta_2}.$$

Then the posterior mean of DLT probability at each dose level is

$$\mathbf{E}[p_j|D_k] = \iint p_j p_k(\beta_1, \beta_2|D_k)d\beta_1 d\beta_2$$

To look for an appropriate dose level for the next trial, our principle is to find the dose level with the DLT probability closest to TTL. Hence the next dose level can be defined as

$$d_{next} = \arg\min_{d_j \in S}(|TTL - \mathbf{E}[p_j|D_k]|).$$

Here S is the set of all permissible choices of dose level.(Babb and Rogatko 2004)

3.1.2 Two-stage likelihood-based CRM

Two-stage likelihood CRM divides the process into two stages. In the first stage, the patients are dosed in single-patient cohorts until the first DLT appears. After the first appearance of DLT, the CRM starts to work on the data based on all the previous trials (first-stage data included).

The stage 2 procedure is similar to the above, but using a maximum likelihood estimation (MLE) to estimate the parameters β_1 and β_2 and calculate the corresponding probability of DLT at each dose level. The estimated parameters based on given data on k dose levels are

$$(\hat{\beta}_1, \hat{\beta}_2) = \arg\max_{(\beta_1, \beta_2)} L(D_k|\beta_1, \beta_2).$$

Here $L(D_k|\beta_1, \beta_2)$ is the same as defined above in the Bayesian setting. Then we can compute the probability to DLT at each dose level under the current MLE of the parameters $p(d_j|\hat{\beta}_1, \hat{\beta}_2)$. Now since we want the next dose level to have the closest probability of DLT to TTL, we're able to define the next dose level by

$$d_{next} = \arg\min_{d_j \in S}(|TTL - p(d_j|\hat{\beta}_1, \hat{\beta}_2)|).$$

Iterate above procedure until the dose level meets the stopping condition. Then the second stage terminates.(Wages, Conaway, and O'Quigley 2013)

3.2 (By Ujjwal Sehrawat)

- Note: content?

3.3 Implementation plan (By Yiwei Ding)

Here we reproduce results in specific figures by fitting a logistic regression model. We run simulations using the CRM with logistic regression, and we derive Figure 6&7. Figure 6 used a one-stage Bayesian approach and Figure 7 used a two-stage likelihood-based approach.

To reproduce Fig. 6, we should use a one-parameter logistic model and placed an exponential prior distribution with a mean of 1. Throughout the trial, calculations were made for the posterior estimates of the likelihood of dose-limiting toxicity (DLT) at each dosage. The upcoming group was then administered the dosage that had an estimated DLT probability most closely aligning with the target toxicity level (TTL).

To reproduce Fig. 7, instead we implemente a two-stage likelihood-based CRM design. We combined a one-parameter power model for the dose-toxicity relationship. During the initial stage, virtual patients were given gradually increasing doses, starting with a dose of 10 ng/kg, which was 1% of the Maximum Tolerable Dose (MTD) in dogs. If a grade 2+ non-DLT adverse event occurred in a patient, two more virtual patients received the same dosage. If none of the trio exhibited a Dose-Limiting Toxicity (DLT), the study continued escalating the dosage in the first stage. Once the first DLT was observed, the second stage, based on the model, was initiated.

A dose skeleton, which was determined after the first DLT (as it wasn't needed during the first phase), was used to establish dose labels for each dose. The likelihood of a DLT at each dose was determined using maximum likelihood methods, and the upcoming patient was assigned the dose that had an estimated DLT likelihood closest to the Target Toxicity Level (TTL). This was under the condition that no dose level that hadn't been tested could be bypassed. Single-patient cohorts were the norm because a low toxicity incidence was anticipated, and each virtual patient was thoroughly observed before the next patient was assigned a dose.

Model Reproduction

- Reproduce results in Figures 6 & 7

```
# R code for model reproduction goes here
```

Simulation Studies

```
# R code for simulation studies goes here
```

Results and Discussion

Answers to the questions

Conclusion

Summary of the report

Reference

Babb, James S, and André Rogatko. 2004. *Bayesian Methods for Cancer Phase i Clinical Trials*. New York: Marcel Dekker.

Brock, Kristian, Lucinda Billingham, Mhairi Copland, Shamyla Siddique, Mirjana Sirovica, and Christina Yap. 2017. "Implementing the EffTox Dose-Finding Design in the Matchpoint Trial." *BMC Medical Research Methodology* 17 (1): 1–15.

Cater, SK. 1972. "Study Design Principles for Clinical Evaluation of New Drugs as Developed by the Chemotherapy Program of the National Cancer Institute." *The Design of Clinical Trials in Cancer Therapy*, 242–89.

Cheung, Ying Kuen. 2011. *Dose Finding by the Continual Reassessment Method*. CRC Press.

Chiuzan, Cody, Jonathan Shtaynberger, Gulam A Manji, Jimmy K Duong, Gary K Schwartz, Anastasia Ivanova, and Shing M Lee. 2017. "Dose-Finding Designs for Trials of Molecularly Targeted Agents and Immunotherapies." *Journal of Biopharmaceutical Statistics* 27 (3): 477–94.

Food, Drug Administration, et al. 2013. "Adaptive Design Clinical Trials for Drugs and Biologics; 2018." *Available at: Fda. Gov/Regulatory-Information/Search-Fda-Guidance-Documents/Adaptive-Design-Clinical-Trials-Drugs-and-Biologics. Accessed [September 12, 2019]*.

Garrett-Mayer, Elizabeth. 2006. "The Continual Reassessment Method for Dose-Finding Studies: A Tutorial." *Clinical Trials* 3 (1): 57–71.

Harrington, Jennifer A, Graham M Wheeler, Michael J Sweeting, Adrian P Mander, and Duncan I Jodrell. 2013. "Adaptive Designs for Dual-Agent Phase i Dose-Escalation Studies." *Nature Reviews Clinical Oncology* 10 (5): 277–88.

Iasonos, Alexia, Nolan A Wages, Mark R Conaway, Ken Cheung, Ying Yuan, and John O'Quigley. 2016. "Dimension of Model Parameter Space and Operating Characteristics in Adaptive Dose-Finding Studies." *Statistics in Medicine* 35 (21): 3760–75.

Iasonos, Alexia, Andrew S Wilton, Elyn R Riedel, Venkatraman E Seshan, and David R Spriggs. 2008. "A Comprehensive Comparison of the Continual Reassessment Method to the Standard 3+ 3 Dose Escalation Scheme in Phase i Dose-Finding Studies." *Clinical Trials* 5 (5): 465–77.

Jaki, Thomas, Sally Clive, and Christopher J Weir. 2013. "Principles of Dose Finding Studies in Cancer: A Comparison of Trial Designs." *Cancer Chemotherapy and Pharmacology* 71: 1107–14.

Le Tourneau, Christophe, Hui K Gan, Albiruni RA Razak, and Xavier Paoletti. 2012. "Efficiency of New Dose Escalation Designs in Dose-Finding Phase i Trials of Molecularly Targeted Agents." *PloS One* 7 (12): e51039.

Le Tourneau, Christophe, J Jack Lee, and Lillian L Siu. 2009. "Dose Escalation Methods in Phase i Cancer Clinical Trials." *JNCI: Journal of the National Cancer Institute* 101 (10): 708–20.

Møller, Susanne. 1995. "An Extension of the Continual Reassessment Methods Using a Preliminary up-and-down Design in a Dose Finding Study in Cancer Patients, in Order to Investigate a Greater Range of Doses." *Statistics in Medicine* 14 (9): 911–22.

O'Quigley, John. 1999. "Another Look at Two Phase i Clinical Trial Designs." *Statistics in Medicine* 18 (20): 2683–90.

O'Quigley, John, Alexia Iasonos, and Björn Bornkamp. 2017. *Handbook of Methods for Designing, Monitoring, and Analyzing Dose-Finding Trials*. CRC press.

O'Quigley, John, Margaret Pepe, and Lloyd Fisher. 1990. "Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer." *Biometrics*, 33–48.

O'Quigley, John, and S Zohar. 2006. "Experimental Designs for Phase i and Phase i/II Dose-Finding Studies." *British Journal of Cancer* 94 (5): 609–13.

Onar, Arzu, Mehmet Kocak, and James M Boyett. 2009. "Continual Reassessment Method Vs. Traditional Empirically Based Design: Modifications Motivated by Phase i Trials in Pediatric Oncology by the Pediatric Brain Tumor Consortium." *Journal of Biopharmaceutical Statistics* 19 (3): 437–55.

Onar-Thomas, Arzu, and Zang Xiong. 2010. "A Simulation-Based Comparison of the Traditional Method, Rolling-6 Design and a Frequentist Version of the Continual Reassessment Method with Special Attention to Trial Duration in Pediatric Phase i Oncology Trials." *Contemporary Clinical Trials* 31 (3): 259–70.

Penel, Nicolas, and Andrew Kramar. 2012. "What Does a Modified-Fibonacci Dose-Escalation Actually Correspond To?" *BMC Medical Research Methodology* 12 (1): 1–5.

Rogatko, André, David Schoeneck, William Jonas, Mourad Tighiouart, Fadlo R Khuri, and Alan Porter. 2007. "Translation of Innovative Designs into Phase i Trials." *Journal of Clinical Oncology* 25 (31): 4982–86.

Thall, PF, and S-J Lee. 2003. "Practical Model-Based Dose-Finding in Phase i Clinical Trials: Methods Based on Toxicity." *International Journal of Gynecologic Cancer* 13 (3).

Wages, Nolan A, Mark R Conaway, and John O'Quigley. 2013. "Performance of Two-Stage Continual Reassessment Method Relative to an Optimal Benchmark." *Clinical Trials* 10 (6): 862–75.

Wheeler, Graham M, Adrian P Mander, Alun Bedding, Kristian Brock, Victoria Cornelius, Andrew P Grieve, Thomas Jaki, et al. 2019. "How to Design a Dose-Finding Study Using the Continual Reassessment Method." *BMC Medical Research Methodology* 19 (1): 1–15.