Continual reassessment and related designs in dose-finding studies

Alexia Iasonos[a,*,†] and **John O'Quigley**[b]

[a]Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York, U.S.A.

[b]INSERM, Université Pierre et Marie Curie, F75005 Paris, France

The workshop on the continual reassessment method and related Phase I designs for dose-finding studies was held at Memorial Sloan Kettering Cancer Center in New York on October 2nd, 2009. The workshop was the result of several discussions between Alexia Iasonos and John O'Quigley. These discussions centered around the methodological advancements in Phase I designs with a particular focus on the continual reassessment method (CRM). The participants of the workshop included applied statisticians who have practical experience in using the CRM in clinical trials, an experience gained both in academic and cancer centers, as well as in the pharmaceutical industry in the United States, Japan and Europe. The majority of the participants have participated in Institutional Review Boards and have been involved in discussions with the FDA and European Medicines Agency. One goal of the workshop was to provide a forum to share the lessons learned from finished trials and to discuss the operational and logistical issues, as well as, methodological advancements in adaptive Phase I clinical trial design.

Alongside the discussion focused on the situations that call for improved statistical approaches, there was concern that, for the simplest of situations, designs that offer very clear improvement over the standard 3+3 design (SD) are still not widely adopted. Rogatko [1] points out that the majority of Phase I trials continue to use the standard design and this despite the fact that leaders in the field have, for several years, called for the standard design to be abandoned in favor of more efficient and more ethical designs [2-7]. We suspect that one reason for the continued use of the standard design is that investigators are hesitant to employ a complicated algorithm in order to assign a dose to a patient. Support for the standard design is often expressed in terms which are vague, such as 'it's been around and in use for quite a while,' 'it's very simple to understand,' 'it's very quick to set up'. Such arguments though, would hold very little sway if we were discussing a Phase III comparative study. It is generally agreed that randomized Phase III trials are the gold standard to answer the question whether drug A is better than drug B, and Phase III trials will often use mathematical models for a drug/treatment assignment, in particular if the design is a complex one. The statistical underpinnings of Phase III trials are not trivial, sample size calculations and randomization procedures are not setup by non-statisticians, and we certainly do not use designs that have been used in the past if they cannot answer the question of interest. Similarly, Phase I dose-finding trials should be following the same principles and ethical code like any other study in human subjects and should utilize statistical designs that can adequately address the primary objective, which is to find the maximum tolerated dose (MTD).

*Correspondence to: Alexia Iasonos, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 307 E. 63rd Street, 3 Floor, New York, NY 10021, U.S.A.. †*iasonosa@mskcc.org*.

One of the less vague, and more compelling, arguments advanced against the newer designs such as the CRM is that they take longer, sometimes much longer, to complete. This is a misconception and indeed, in many cases the opposite is true. This is very easily seen to be the case. In fact, it is not possible to proceed more quickly than a CRM design since patients can be included as and when they become available for study. This is not the case for the SD which is, at best, as rapid as a CRM design but, generally, slower since, before proceeding to the next dose allocation, all 3 patients in the most current cohort must have been observed for toxicity. The probable source of this misconception is the assumption that it is necessary to restrict CRM inclusion to one patient at a time and await the outcome before proceeding to the next inclusion. However, not only was this not a requirement of the method as originally described in O'Quigley *et al.* [7] – a whole section, Section 3 of that paper, was devoted to grouped inclusions as well as the inclusion of patients before the results of previously entered patients had become available – but, in order to respect the underlying rule of the CRM, we should always include patients at our best current estimate of the MTD. Clearly this does not require that we wait for any particular outcomes. A decision to wait would be a clinical one and would apply to any and all designs, that is to say any such delay is not a result of the design itself. Goodman *et al.* [8] carried out an exhaustive study of CRM designs using grouped inclusions and concluded that the operating characteristics, when compared to one by one inclusion, were very similar. Thall *et al.* [5] provided specific guidelines for accelerating trial duration and Iasonos *et al.* [9] showed that the CRM is both faster and requires a smaller sample size compared to the SD as the number of tested dose levels increases.

Another common misconception is that model-based designs put patients at higher risk than those included in studies using the standard design. First, this requires a particularly narrow definition of risk in that under-dosing, regardless of by how much, is not taken to be any kind of risk. However, even under this definition, in which only the allocation of patients to levels higher than the MTD is taken to correspond to an increase in risk, it is not true that the model-based design will over-dose more than the standard design. One of the advantages of the CRM is the fact that it incorporates a tuning parameter for what the investigator considers to be an acceptable rate of toxicity and this can vary while the standard design does not. Theoretical investigations on this question [10, 11] have shown clearly that the probability of over-dosing is greater with the standard design, something which is easily anticipated given the control allowed the investigator by way of design flexibility. This flexibility allows the CRM to target any percentile, in particular one lower than one-third, to employ distance measures which weight in favor of under-dosing rather than over-dosing, as in designs with overdose control [12], and to exploit information on lower grade toxicities as in [13] in finding the MTD. The view that CRM trials will put patients at higher risk is based on a misinterpretation of data [14]. When allocating patients to a dose level on the basis of a fitted CRM model, it is impossible to escalate following a toxicity [14, 15].

Among the aims of a clinical trial is the aim to minimize harms and maximize benefits. This requires the clinical team to use the best possible research design, even if this design is not trivial and requires statistical programming as well as more complex logistical issues involved around the dose assignment. In the Phase I context, the patients know that the probability of benefit is small, and the risk of serious potentially life-threatening adverse effect is large. Nonetheless, Phase I patients do expect to benefit otherwise they would most likely not consent. Indeed, the notion of beneficence is one of the founding Belmont principles and all studies on human subjects are required to adhere to the Belmont principles. Phase I patients on trials that follow the SD may not obtain any benefit, especially if they are among the earlier inclusions, since they have a high probability of receiving a drug at sub-therapeutic levels even if the drug turns out to be efficacious in later testing. It has been shown by many authors that the SD is sub-optimal because it starts at the

lowest level and it will treat more patients than is reasonable at sub-therapeutic levels [9, 16, 17]. This is a result of the method's lack of flexibility to treat patients at the level which appears to be the best in the light of all current knowledge, knowledge obtained from within and without the study. As experimentation proceeds, this knowledge will change but, at every step, if we are to respect the Belmont principles, then it is very difficult to justify a design such as the standard design which, consciously, will include many patients at sub-therapeutic levels.

A number of model-based Phase I designs have been described in the statistical literature and they are all generally superior to the SD. Once we go beyond the simplest of cases, with the aim of developing designs that more accurately reflect the actual clinical context, we must recognize that the SD has no facility to do this. The SD works with a single binary outcome of toxicity measured in a homogeneous group of patients when the orderings between the dose levels are known. For more complex situations, such as dose-finding studies in two distinct patient cohorts, or studies involving drug-combinations the SD offers no potential to adapt to these more complex circumstances and, in the presence of different groups, different schedules or different treatment combinations, the recommended dose from a study based on the SD is often amended in Phase II trials. Only model-based designs enable us to address explicitly, right from the outset, such added complexity which, in practice, is the rule rather than the exception.

The focus of the workshop was on these more complex situations. Questions of patient heterogeneity, of within patient dose escalation, of multidrug studies, of ordinal toxicity grading as an outcome, of bridging studies, of combining information from several studies, of tackling the issues of different treatment schedules and of dose-finding studies in which the ordering between the dose levels was not fully known—the multidrug problem in particular—were all discussed. In conjunction with that discussion, focused on practical clinical issues, we also considered statistical questions such as ways to improve the performance of some of the basic designs and ways to estimate the dose–toxicity curve overall following any completed study.

Following a general introduction and overview of the field, given by Alexia Iasonos, John O'Quigley gave a talk presenting simple extensions of model-based designs. These extensions amount to appealing to more than a single model. Yin and Yuan [18] used a similar idea in an effort to gain robustness and reduce sensitivity to model choice in CRM studies. The idea here was not so much to gain added robustness, but more to gain added flexibility in dealing with more complex situations that can arise in practice. The greater flexibility enables us to take on board many different kinds of added complexity. Examples include extended models to deal with subject heterogeneity, extended models to take account of different treatment schedules and extended models to tackle the problem of partial ordering. The added flexibility is controlled very readily by the investigator so that, for instance, it is quite straightforward to limit potential models in the light of information we have. As an example, consider the problem of heterogeneity limited to two groups—for illustrative purposes we could imagine these to be a group of heavily pre-treated patients as opposed to a group with a relatively lighter pre-treatment. We may wish to allow for the possibility of different MTDs while, at the same time, not allowing the difference to be greater than, say, one or two levels. This would make sense since the definition itself (heavy versus less heavy pre-treatment) is not a sharp one. The MTDs may differ but not realistically by more than one or two levels. We may also know that if there is any difference between the MTDs, then it must be in a certain direction, as an example it must be that the heavily pre-treated group have an MTD no greater than that for the other group. It may be surprising but the inclusion of such simple additional information can greatly improve accuracy of estimation. The inferential problem is quite straightforward if we can

make some working assumptions about the structure of the models. We can appeal to established results on Bayesian model choice. The difficulty arises as a result of the under-parameterization of the models so that the 'true' model may not find itself among our class of available models. Simulations are very encouraging but more theoretical work is needed to investigate questions of Bayesian model choice for these kinds of models.

Elizabeth Garrett-Mayer presented the proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading. Her co-authors extended the CRM to include ordinal toxicity outcomes as specified by Common Toxicity Criteria using the proportional odds model and compared the results with the dichotomous CRM. A sensitivity analysis of the new design compared various target dose-limiting toxicity rates, sample sizes and cohort sizes. This design was also assessed under various dose–toxicity relationship models including proportional odds model as well as those that violate the proportional odds assumption. A simulation study showed that the proportional odds CRM performs as well as the dichotomous CRM on all criteria compared (including safety criteria such as percentage of patients treated at highly toxic or sub-optimal dose levels) and with improved estimation of the MTD in certain situations. These findings suggest that it could be beneficial to incorporate ordinal toxicity endpoints into Phase I trial designs. Having several parameters to estimate in studies with such small sample sizes is a challenge and the limitations of the approach were discussed.

In a Bayesian setting, Ying Kuen Cheung introduced the concept of a least informative prior variance using a normal prior distribution and proposed approaches to jointly calibrate the prior variance and the initial guesses of the probability of toxicity at each dose given the target probability of toxicity, the number of dose levels, the sample size and the functional form of the dose–toxicity model. Cheung and Lee addressed the calibration of the prior distribution of the parameter of interest for the Bayesian CRM which has not been previously addressed in the literature. While the results of the new approaches are very similar to the previously proposed methods given the appropriate vague prior chosen by these methods, the new approaches yield a smaller indifference interval. Given the small sample sizes that we nearly always face in this context, this added precision can translate itself as a sharper estimate of the MTD. In addition, the systematic approach to determining the model skeletons, as seen by other investigators, will help not just with precision but also with the problems of robustness.

Ying Yuan presented a hybrid Bayes factor design for Phase I oncology clinical trials. The idea is to obtain a balance between robustness and efficiency by appealing to a careful combination of parametric and nonparametric models. When the observed data at the current dose do not contain enough information to make an informative decision, the authors' suggestion is to appeal to a parametric dose–toxicity curve to borrow strength across all the doses under study to guide dose assignment. As a hybrid of nonparametric and parametric methods, the proposed approach inherits the robustness of the nonparametric methods and the efficiency of the model-based methods. Yuan and Yin examined the properties of the hybrid Bayes factor design through extensive simulation studies, and also compared the new method with the original CRM. The simulation results show that the proposed design is competitive with other available dose-finding methods, and it is more robust than parametric models while it is more efficient than model-free methods.

Satoshi Morita presented an application of the Bayesian CRM to a Phase I dose-finding study in Japanese patients with advanced breast cancer using an informative prior elicited from clinical investigators. One of the principal advantages of applying a Bayesian CRM is the utilization of all available prior information to estimate the recommended dose through prior distributions that are assumed for model parameters representing the dose–toxicity

relationship. In some settings, it may be appropriate to use an informative prior that reflects the accurate and comprehensive previous knowledge of clinical investigators. An earlier completed study in a different country can provide some basis for the construction of prior knowledge. At the same time, because of potentially important differences in treatment tolerability between Caucasians and Japanese there is some need to give careful weighting to the different sources of information. It would not be desirable for the prior data to dominate decision making. In the long run of course, it is the accumulating data that dominate decisions as the amount of observed data in Japanese patients increases. The difficult issue is to gauge how much influence the prior data should have early in the study when only a few Japanese patients have been included. Dr Morita discussed the relative strength of the prior using a recently proposed method to compute a prior effective sample size and used this technique to moderate the effective input of the prior.

Sarah Zohar presented an approach to meta-analysis of dose-finding studies. For cytotoxic clinical trials in oncology, it is not unusual to carry out several Phase I studies on a new molecule or procedure. For instance, the molecule Sorafenib (which inhibits particular tyrosine kinase enzymes in several cancers) was used alone in five published clinical trials. Each clinical trial was conducted separately in different indications and the resulting data were pooled for illustration. In order to justify this, it was argued that the outcome of toxicity may be unrelated or weakly related to disease itself. Integrating information across several Phase I trials may lead to improved inference on the dose level, or levels, corresponding to the MTD. If no assumptions are made, a pooled analysis will perform no less well than the several separate analyses. If some slightly stronger assumptions are made, then there is the potential for useful gains. Dr Zohar presented a method that retrospectively analyzes data obtained according to a dynamic sequential design since a problem with a meta-analysis is that it requires retrospective analysis of data that were obtained in a sequential fashion.

Alexia Iasonos proposed a method to analyze completed phase I trials and possibly confirm or amend the recommended Phase II dose, based on constrained maximum likelihood estimation (CMLE). Dr Iasonos presented a comparison of CRM, retrospective-CRM (as described in [19]), isotonic regression and CMLE in analyzing simulated trials that had followed the standard design. The authors hypothesize that a retrospective analysis would suggest an MTD that is more accurate than the one obtained by the standard design, especially in cases when deviation from the original trial design occurs, since the rules for determining MTD are no longer applicable. The results show that CMLE more accurately selects the true MTD than the standard design, and is better or comparable to isotonic regression and retrospective CRM. Confidence intervals around the toxicity probabilities at each dose level can be estimated using the cumulative toxicity data and decisions whether the dose should be amended could take these into account.

All the above papers appear in this issue. However, there were additional presentations that have been published or being published elsewhere. These included the following topics: (1) Thomas Braun from University of Michigan presented two models in the CRM framework that incorporate patient heterogeneity by examining how random effects can be used in adaptive Phase I trial designs to account for patient heterogeneity. He presented the simulation results of comparing these models relative to models that ignore heterogeneity. (2) Peter Thall from University of Texas, M.D. Anderson Cancer Center presented a Bayesian outcome-adaptive design for choosing the optimal dose pair of two agents used in combination in a Phase I/II clinical trial [20]. Patient outcome was characterized as two ordinal variables accounting for toxicity and treatment efficacy. The method was illustrated by a trial of a chemotherapeutic agent and a biologic agent for treatment of bladder cancer. (3) Glen Laird from Novartis Pharmaceuticals Corporation reported the group's experiences

with the one-parameter CRM and a comparison with conclusions obtained from a two-parameter CRM when applied in three clinical trials. The group showed that the concern of over-dosing patients is unwarranted if (a) a reasonably flexible model is used to estimate the dose–toxicity curve and (b) inferential summaries for DLT rates that address clinical safety concerns (e.g. over-dose probabilities) are taken into account for dose recommendations. (4) Xiaobu Ye from Johns Hopkins University School of Medicine shared the experience of using modified CRM method in dose-finding trials conducted through the new approaches to brain tumor therapy consortium. Dr Ye demonstrated the software and implementation of the modified CRM method, and also discussed challenges with regard to the new anticancer drug paradigm of molecularly targeted agents.

References

1. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. Journal of Clinical Oncology. 2007; 25(31):4982–4986. [PubMed: 17971597]

2. Ahn C. An evaluation of phase I cancer clinical trial designs. Statistics in Medicine. 1998; 17(14): 1537–1549. [PubMed: 9699228]

3. Piantadosi S, Liu G. Improved designs for dose escalation studies using pharmacokinetic measurements. Statistics in Medicine. 1996; 15(15):1605–1618. [PubMed: 8858785]

4. Whitehead J, Williamson D. Bayesian decision procedures based on logistic regression models for dose-finding studies. Journal of Biopharmaceutical Statistics. 1998; 8:445–467. [PubMed: 9741859]

5. Thall PF, Lee JJ, Tseng CH, Estey EH. Accrual strategies for phase I trials with delayed patient outcome. Statistics in Medicine. 1999; 18(10):1155–1169. [PubMed: 10363337]

6. Storer B. An evaluation of phase I clinical trial designs in the continuous dose-response setting. Statistics in Medicine. 2001; 20(16):2399–2408. [PubMed: 11512130]

7. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics. 1990; 46(1):33–48. [PubMed: 2350571]

8. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. Statistics in Medicine. 1995; 14(11):1149–1161. [PubMed: 7667557]

9. Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in Phase I dose-finding studies. Clinical Trials. 2008; 5:465–477. [PubMed: 18827039]

10. O'Quigley J. Another look at two phase I clinical trial designs (with commentary). Statistics in Medicine. 1999; 18:2683–2692. [PubMed: 10521858]

11. Reiner E, Paoletti X, O'Quigley J. Operating characteristics of the standard phase 1 clinical trial design. Computational Statistics and Data Analysis. 1999; 30:303–315.

12. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. Statistics in Medicine. 1998; 17(10):1103–1120. [PubMed: 9618772]

13. Iasonos A, Zohar S, O'Quigley J. Two-stage continual reassessment method that determines dose-escalation based on individual toxicity grades. Society of Clinical Trials 31st Annual Meeting. 2010 (abstract P94).

14. O'Quigley J. Theoretical study of the continual reassessment method. Journal of Statistical Planning and Inference. 2006; 136:1765–1780.

15. Cheung YK. Coherence principles in dose-finding studies. Biometrika. 2005; 92(4):863–873.

16. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. Clinical Trials. 2006; 3(1):57–71. [PubMed: 16539090]

17. O'Quigley J, Zohar S. Experimental designs for phase I and phase I/II dose-finding studies. British Journal of Cancer. 2006; 94:609–613. [PubMed: 16434987]

18. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. Journal of the American Statistical Association. 2009; 104(487):954–968.

19. O'Quigley J. Retrospective analysis of sequential dose-finding designs. Biometrics. 2005; 61:749–756. [PubMed: 16135026]

20. Houede N, Thall PF, Nguyen H, Paoletti X, Kramar A. Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. Biometrics. 2010; 66(2):532–540. [PubMed: 19673865]