Response to comments on 'Competing designs for drug combination in phase I dose-finding clinical trials' by G. Yin, R. Lin and N. Wages

Marie-Karelle Riviere, Frédéric Dubois, Sarah Zohar

SIM-12-0373 and sim-14-0573: Response to 'Comments on Competing designs for drug combination in phase I dose-finding clinical trials' by G. Yin, R. Lin and N. Wages

Marie-Karelle Riviere[1,2,*], Frédéric Dubois[2] and Sarah Zohar[1]

1   INSERM, U1138, Equipe 22 Centre de Recherche des Cordeliers, Université Paris 5, Université Paris 6, France
2   IRIS (Institut de Recherches Internationales Servier), Suresnes, France
* Correspondence to: Marie-Karelle Riviere, INSERM, U1138, Equipe 22, Centre de Recherche des Cordeliers, Université Paris 5, Université Paris 6, Paris, France.
E-mail: marie-karelle.riviere@crc.jussieu.fr

As pointed out by G. Yin and R. Lin and N. Wages, designs for drug-combination trials need to be rigorously evaluated. In our paper we performed an extensive sensitivity analysis (with some of the results given in the form of additional material) and for each design we tried to find the best parameter settings for model calibration associated with a high percentage of correct selection (PCS) on average across all scenarios.

Replying to G. Yin and R. Lin concerning the BCOPULA design, we computed this design using the C code directly provided by Y. Yuan [1], (the C code used in the original manuscript describing the method). Since we only removed the stopping rule there could not be any programming issues. When using the same setting as G. Yin and R. Lin (the authors of the comments), we obtained similar results (see Table below).

In their commentary, Yin and Lin did not use the same designs as in our paper, as (1) the working model was different in that the target toxicity was at the highest dose level; (2) the spacing between prior toxicity probabilities was different from ours; and (3) the cut-offs for BCOPULA and BGUMBEL were not the same as those detailed in the manuscript. However, when we ran simulations using exactly the same settings as stated in the correspondence of Yin and Lin, we obtained the following results:

		sc1	sc2	sc3	sc4	sc5	sc6	sc7	sc8	sc9	sc10
Copula	Yin and Lin	57.9	68.5	58.6	84.4	91.6	40.6	49.6	30.4	19.6	29.5
	Riviere et al. cohort size (3,3)	60.4	69.9	59.1	85.6	91.5	43.3	49.4	33.1	16.3	31.9
	Riviere et al. cohort size (1,3)	53.2	67.8	55.5	85.5	88.9	40.4	51.3	25.7	12.8	31.1
Gumbel	Yin and Lin	55.3	70.3	58.6	85.2	91.4	42.2	50.2	27.4	19.5	29.4
	Riviere et al. cohort size (3,3)	58.7	71.5	58.1	87.2	90.1	45.3	51.9	33.9	22.8	28.7
	Riviere et al. cohort size (1,3)	53.9	69.5	54.9	87.2	88.9	41.2	52.3	29.8	12.8	30.9

We ran simulations with a cohort size of 3 throughout the trial, and then a cohort size of 1 for the start-up phase and a cohort size of 3 for the model-based phase, as it seems this is what Yin and Lin chose in the commentary. We can observe that our results are similar to those of Yin and Lin.

Yin and Lin argue that our working model for the combination of agents is wrong. Firstly, some clinical trials are performed where only one agent is studied in phase I with the other agent being a new molecule for which no prior information is available. In this case, there is no reason to limit the settings of the working model. Secondly, the spacing between the prior toxicity probabilities in our choice of working model is not the same as that of Yin and Lin, and this choice also impacts on performance. Finally, our choice was mainly based on the sensitivity analyses we performed to assess the influence of several working models and their impact on performance. Moving the targeted toxicity from the lowest dose to the highest dose increases PCS in some scenarios and decreases them in others. Choosing the working model with the target ranked in the middle was a good compromise. Of course, this choice is subject to debate (as with any other choice). A best working model could be chosen for each scenario but there was no specific working model that gave the best performance in all cases. In addition, the choice of the working model should also be made according to the information available about each agent before beginning a clinical trial. In our paper, the choices of the working models for the studied designs were performed for the first eight scenarios, before scenarios 9 and 10 were added during the manuscript's review process. It showed that, depending on the retained scenarios and the clinical setting, the working model should be re-calibrated. This was discussed in detail in our paper.

Replying to N. Wages, as stated in our paper, we used the POCRM design starting at the lowest dose level, based on current practice in Phase I oncology clinical trial. We chose to study the Bayesian version of POCRM [2] by analogy to the CRM that is used in most published clinical trials under the Bayesian framework. In [2], a start-up phase was proposed which was different from the one mentioned by Wages in this commentary and the dose allocation did not begin at the lowest dose level. The authors proposed in [2] a start-up phase where at each step the ordering selected was sampled randomly with the weights of posterior probabilities. This start-up phase was not used by the authors in [2] in the simulation study and we chose to proceed as they did. Indeed, in the MLE POCRM paper [3] published soon after, a start-up phase, necessary in the frequentist context, was proposed. As for the Likelihood CRM, the exploration begins at the lowest dose level, escalates by diagonal zones, and ends at the first observation of a DLT. We agree that this start-up phase could be used in the Bayesian POCRM version to begin safely at the lowest dose level. This point was discussed in our paper, we agree that a safe start-up phase is required for the Bayesian POCRM [2], we also have proposed one.

However, our choice of starting dose for comparison purpose did not impact the PCS of the Bayesian POCRM (see results table below)

	sc1	sc2	sc3	sc4	sc5	sc6	sc7	sc8
POCRM (lowest combination)	72.7	64.4	72.5	73.8	81.8	49.1	47.7	55.1
POCRM (prior MTD)	73.1	61.0	75.6	73.7	85.3	50.6	50.1	54.7

The results presented in our paper reflect the PCS of the Bayesian POCRM [2] independently of the starting dose chosen. Regarding the comment concerning the R package [4] mentioned by Wages, this package does not allow implementing the Bayesian POCRM. Moreover, (1) in particular cases there can

be large difference in performance between the two versions of the POCRM; and (2) the package was not flexible enough for the choice retained in our comparative analysis, for example the size of the start-up phase, as well as other parameters, cannot be changed.

Concerning the choice and the number of orderings we have performed our own sensitivity analysis, as the guidelines proposed in [5] were not available on the public domain until late in the review process of our manuscript. However, now that these guidelines are available, we recommend using them. Our sensitivity analysis included several settings: choice of orderings, number of orderings and working model (given in additional material). In the following figure, orderings (a), (b) and (c) correspond to the three orderings, and orderings (a) to (g) to the seven orderings used in the sensitivity analysis on the number of orderings given in the additional material.
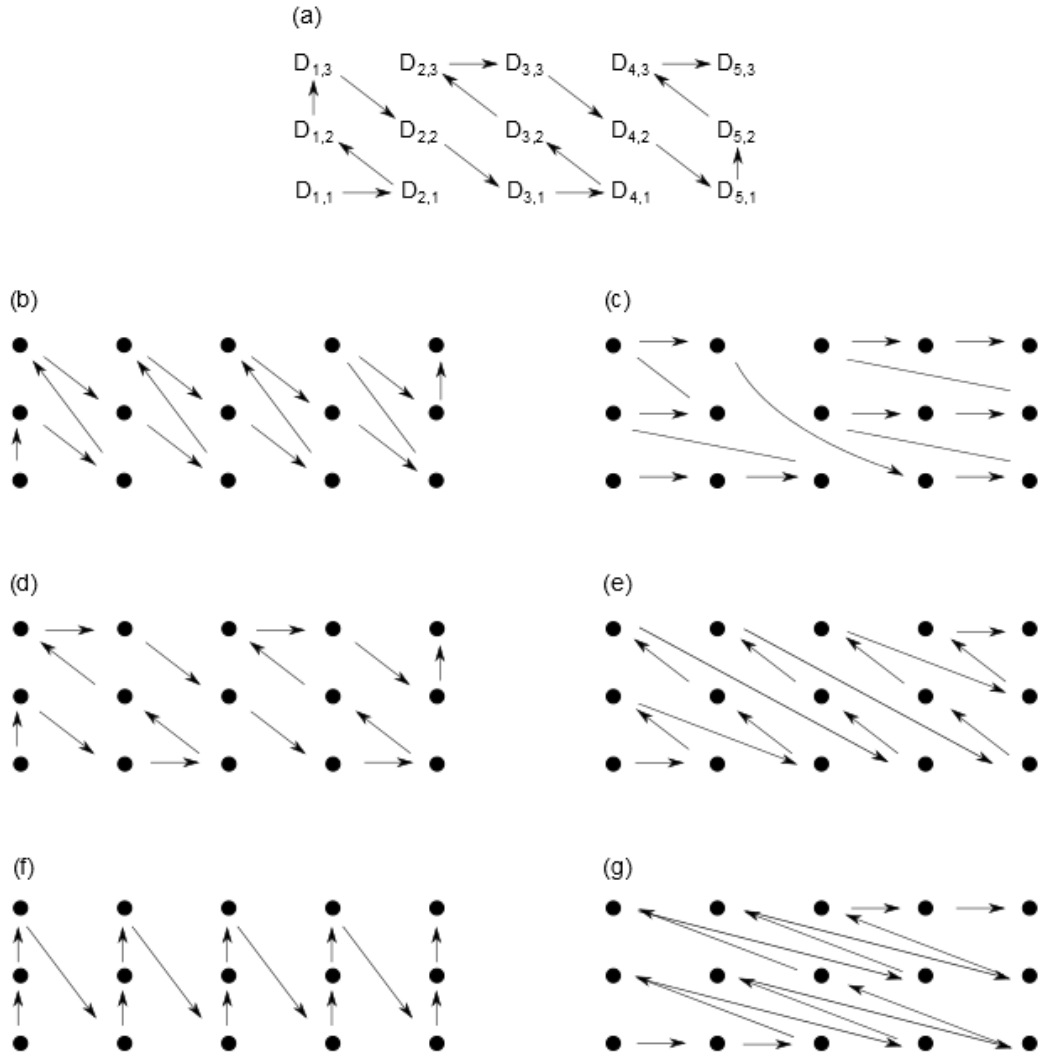
Then in the "22" and "200 orderings" used in the additional material, the previous 7 orderings were included and the remaining orderings to select were sampled in the set of all possible orderings satisfying the partial order (composed of 6006 possible orderings). This sensitivity analysis on the

number of orderings performed on the first eight scenarios led to our choice of three orderings (sc9 and sc10 were added lately during the review as mentioned above).

We believe that, in general, it is always possible to find some scenarios that tumble the performance of a method. Scenarios 9 and 10, that are peculiar (as they do not respect the "zoning" assumption often supposed in combination designs as well as in the start-up phase proposed by Wages [3]), were added to highlight the benefit of multi-dimensional designs vs one-dimensional designs. Therefore, we did not observe the lack of information due to the 3 orderings compare to 6 orderings. The use of 6 orderings seems to be an efficient setting for the POCRM as it increases the performance on the last two scenarios. Nevertheless, as we mentioned above, we believe that it is always possible to find some scenarios that tumble the performance of a method. For example, if we choose the following scenario:

0.30   0.45   0.50   0.60   0.75

0.05   0.10   0.15   0.45   0.50

0.01   0.02   0.05   0.08   0.15

then, the PCS for the POCRM given the six orderings proposed by Wages is only equal to 3.4%. Therefore, in general, the performances of all model-based methods seem to remain comparable.

Yin and Lin, and Wages have raised an important question concerning model tuning or calibration. Calibrating such complex models requires rigorous simulation studies. As an example, the model-based continual reassessment method (CRM) using a 1-parameter dose-toxicity model has been the subject of many papers concerning its calibration. Numerous authors have tried to highlight how to calibrate the CRM before and during a dose finding clinical trial, including for example Cheung and Chappell [6], Lee and Cheung [7, 8], Zhang et al. [9], Yuan and Yin [10], Daimon et al. [11] among others. For the CRM the choice of the working model and the prior (when the Bayesian framework is used) considering only a one-parameter model has been the subject of much debate and there is still no consensus. In combination model-based designs, the models are more complex and there are more parameters to calibrate. The parameter settings used in a methodological manuscript of a proposed model are not necessarily the ones that are the most suitable for a specific indication and disease. Therefore, before the beginning of each new dose-finding clinical trial, an extensive simulation study should be performed in order to calibrate the design.

In conclusion, based on the assumptions we have made, our numerical results are valid when choosing the same setting values. The choice of the working model and calibration is important but much more complex in the case of combination studies. Given a set of fixed scenarios, we can always tune the design to make it better, but this may not work in all situations and all scenarios. Indeed, for model-based combination methods, this aspect still needs to be studied as it was done for the CRM. In our manuscript, we did not favor any design over the others; this should be left for discussion between the members of the team involved in the planning of the clinical trial. The choice of the design should be made according to the clinical setting, the indication, the disease, the knowledge about each agent prior to trial onset and an extensive sensitivity analysis.

References

1. Yuan Y, Yin G. Robust EM Continual Reassessment Method in Oncology Dose Finding. Journal of the American Statistical Association 2011; 106(495):818–831.

2. Wages NA, Conaway MR, O'Quigley J. Continual reassessment method for partial ordering. Biometrics 2011; 67:1555–1563.

3. Wages NA, Conaway MR, O'Quigley J. Dose finding design for multi-drug combinations. Clinical Trials 2011; 8:380–389.

4. Wages NA, Varhegyi N. pocrm: an R package for phase I trials of combinations of agents. Computer Methods & Programs in Biomedicine 2013; 112:211–218.

5. Wages NA, Conaway MR. Specifications of a continual reassessment method design for phase I trials of combined drugs. Pharmaceutical Statistics 2013; 12:217–224.

6. Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the continual reassessment method. Biometrics 2002; 58(3):671–674.

7. Lee SM, Cheung YK. Model calibration in the continual reassessment method. Clinical Trials 2009; 6(3):227–238.

8. Lee SM, Cheung YK. Calibration of prior variance in the Bayesian continual reassessment method. Statistics in Medicine 2011; 30(17):2081–2089.

9. Zhang J, Braun TM, Taylor JM. Adaptive prior variance calibration in the Bayesian continual reassessment method. Statistics in Medicine 2013; 32(13):2221–2234.

10. Yin G, Yuan Y. Bayesian dose finding in oncology for drug combinations by copula regression. Journal of the Royal Statistical Society 2009; 58:211–224.

11. Daimon T, Zohar S, O'Quigley J. Posterior maximization and averaging for Bayesian working model choice in the continual reassessment method. Statistics in Medicine 2011; 30(13):1563–73.