Avenue E. Mounier 83/11
1200 Brussels
Belgium
Tel: +32 2 774 1611
Email: eortc@eortc.be
www.eortc.org

Trial Design – Phase I Trials

ST-001-WIN-01
Version: 2
ALWAYS REFER TO INTRANET TO CHECK THE VALIDITY OF THIS DOCUMENT

Author *Associate Head of Statistics Department* Laurence Collette	Signature:	Date: 07 Dec 15
Approved and authorized by *Methodology Director* Jan Bogaerts	Signature:	Date: 7 December 2015

1 PURPOSE

The current work instruction details various aspects of the statistical design of EORTC Phase I trials (refer to glossary) as a function of the objectives of the study.

2 DEFINITIONS

- **Maximum Tolerated Dose (MTD)**: the EORTC employed MTD definition is the lowest dose that produces an unacceptable rate of DLTs, generally 2 out of 3 or 2 out of 6 patients. (The recommended phase II dose is then the dose level just below the MTD).
- **Dose Limiting Toxicity (DLT)**: specific toxicities observed over a defined assessment period, at a grade (and possibly minimum duration) that is regarded as not tolerable . Typically for cytotoxic agents the assessment period is cycle 1 and the DLTs are any grade 4 hematological toxicity or any grade 3/4 non-hematological toxicity.
- **Recommended Phase II dose (RPIID)**: Recommended dose classically defined for cytotoxic agents based on the 3+3 design as the dose level dose level at which less than one third of patients experience severe toxicities (dose limiting toxicities [DLTs]) during cycle 1. For a model based design with a target toxicity rate θ (usually 20%) it will be the dose that produces the toxicity rate below and closest to the target.

3 INSTRUCTION

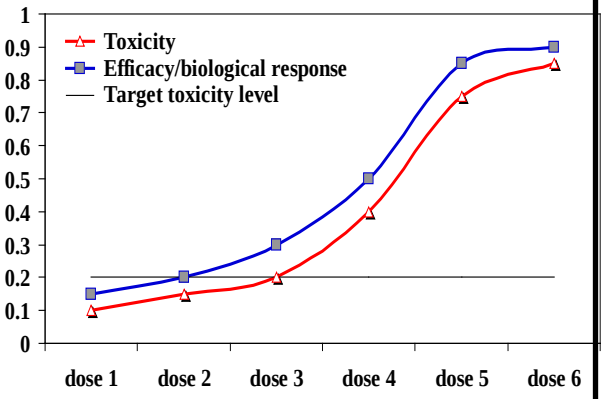
3.1 Introduction

The design of phase I studies was originally thought of for the investigation of cytotoxic drugs. Such studies are designed based on the assumption of increasing dose-effect relationships, both for efficacy and for adverse events. Designs presented in 3.2 below for cytotoxic drug studies should not be blindly applied in other situations, for example when investigating molecularly targeted agents for which this assumption may not hold. For such situation, please refer to section 3.3 for further description.

For reference, the below graphs are provided as hypothetical situations of dose-toxicity and dose-efficacy curves for cytotoxic vs cytostatic agents.

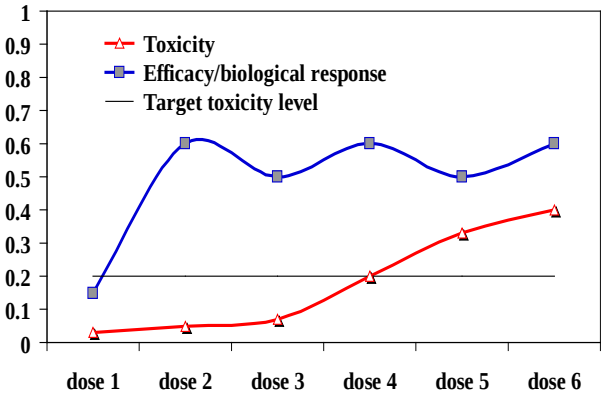
Figure 1

*Dose-toxicity and efficacy curves
Cytotoxic agents



- Higher toxicity associated with higher efficacy.
- Need to balance toxicity with efficacy.

Dose-toxicity and efficacy curves
Cytostatic/biologic agents



- Shallower dose-toxicity curve
- Non-monotonically increasing dose-efficacy curve
- Dose selection driven by safety and biological effect
- Proof-of-principle of biologic effect in early trials

It is the Statistician's task to ensure that the eligibility criteria are restricted to elements that ensure appropriate assessment of the primary endpoint, and appropriate delineation of the population of interest.

Therefore, Phase I trials are generally open to all tumor types, however when testing targeted agents, eligibility may sometimes be restricted upfront to patients whose tumor presents the mutation of interest.

3.2 Phase I trials for drugs for which a increasing dose-toxicity and dose-activity relationship may be assumed

3.2.1 Objectives

Phase I trials aim to screen for toxicity as the dose is gradually escalated. The fundamental assumption is that both the anti-tumor effect and the toxicity increase as the dose is increased (therefore, if there is no expectation of toxicity increasing with dose the designs proposed below are fundamentally flawed). Safety (toxicity, pharmacology) is the limiting factor in these studies that aim to increase the dose to the maximum possible, with the hope that efficacy will be optimized with increased dose.

The usual objectives of such phase I studies are to:

- Primary: to propose a recommended phase II dose (RPIID)

By

a) establishing the Maximum Tolerated Dose (MTD) for the new drug for a given schedule via a given route of administration and propose a dose for phase II evaluation in man that is an acceptable tradeoff between potential activity and toxicity.

b) indentifying the Dose Limiting Toxicity (DLT), both qualitatively (which organ system is involved) and quantitatively (predictability, extent, duration and reversibility).

- Secondary: to determine the pharmacokinetic and the pharmaco-dynamic profile of the drug and to document possible anti-tumor activity

3.2.2 Endpoints

3.2.2.1 Primary endpoint

The primary endpoint in phase I trials is thus generally the adverse reactions related to the drug. More specifically, the establishment of the MTD and the identification of the DLT are the outcomes of interest in these statistical designs.

Together with the Study coordinator (SC) and the Clinical Research Physician (CRP), the Statistician ensures that the definition of what constitutes a DLT is unambiguous and verifiable from the database and the study protocol. The period during which the defined events is counted as DLT (e.g. during cycle 1, first 2 cycles …) must be specified and the requirements for patient evaluability should be clearly stipulated. In particular, how patients that do not complete the period of interest for other reasons than encountering a DLT will be treated in the interpretation must be clear

Of note, the EORTC employed MTD definition is the lowest dose that produces an unacceptable rate of DLTs, generally 2 out of 3 or 2 out of 6 patients. The RPII is then the dose level just below the MTD.

3.2.2.2 Secondary endpoints

Secondary endpoints can be clinical response and pharmacokinetic data, with the ultimate objective of correlating pharmaco-kinetics/pharmacodynamics (PK/PD) to drug activity.

3.2.3 Statistical design

The statistical design intends to identify the MTD based on the occurrence of DLTs. In most trials, only the evaluation of the toxicity during the first cycle is used to establish the MTD. Increasing dose levels are sequentially investigated in successive patients. After each inclusion of cohorts of 1 or more patients, the next level to be administered is assessed.

Phase I designs are not always conducted by strict rules. The main limitation is the sample size (between 12 and 30 patients). The statistical design should serve as a guide to allocate the doses and to make the fullest use of the information contained in the data.

During the trial, there should be a continuous discussion between the Statistician, the Clinical Research Physician and the Study Coordinator.

3.2.3.1 Starting dose and dose increment

The starting dose is based on pre-clinical studies. Various methods for selecting a safe starting dose are used. For example in EORTC studies, one-tenth of the LD10 in mice (dose that is lethal in 10% of the mice), expressed as mg/m2, is generally used as the starting dose for phase I trials, provided that this dose is not toxic in a second species (normally the rat).

The successive dose levels to be investigated are based on a sequence of increments that are generally independent of the design used to allocate patients to the various dose levels. When no maximum dose is known in advance, only a mechanism to increment the different doses is given. Different rules may be

applied based on the observed toxicity. In case possible early onset of toxicity is considered, a dose level -1 (below the starting dose) can be defined as well.

A classical method of dose escalation (called Fibonacci series or increments) specifies that successive dose levels are increased by 100%, 67%, 50%, 40%, 33%, 33%, 33%... above the preceding dose level. Another possibility is to base the next higher dose level on the toxicity observed at previous levels, for example to increase by 100% when mild toxicity has been reported but by only 50% in case of moderate toxicity.

3.2.3.2 Dose escalation algorithm

The statistical design takes into account the following conflicting operating requirements:

- To minimize the number of under treated patients. Doses far below the MTD are unlikely to have any anti-tumor effect.
- To minimize the number of over treated patients. A too high dose puts the patient at excessive risk of severe toxicity.
- To minimize the sample size since the compound has never been demonstrated to have an anti-tumor effect.

Nevertheless, it should be borne in mind that the dose recommended at the end of the trial will probably not be modified in phase II or phase III studies. During the development of the compound, an erroneous MTD may not even be detected. It is then important not to “sacrifice” the sample size to speed up the drug development.

Several different schemes have been developed.

3.2.3.2.1 Standard algorithmic Method

The standard method (also called the 3 + 3 scheme) is the most widely used technique since it does not require statistical expertise. An initial cohort of 3 patients is given the new compound at the starting dose level and the toxic effects are carefully observed. If no DLT is observed after a predetermined period of observation, the dose is escalated (usually based on the smoothed Fibonacci series) in another cohort of 3 patients and toxicity is again documented. If one DLT is reported among the 3 patients, another 3 patients are treated at that dose level. The dose escalation procedure is continued until 2 out of 3 or 2 out of 6 patients experience DLTs at a given dose, a level of toxicity that is usually considered to be unacceptable. A minimum of 6 patients are to be treated at the dose level to be recommended for phase II studies and kept on treatment to determine chronic or cumulative toxicity. Depending on the case, a larger cohort up to 12 patients can be used as an assurance of not overseeing key toxicities (expansion cohort).

The two main criticisms of this design are the high number of patients included at very low dose levels which have little or no anti-tumor effect and the poor accuracy of the final recommended dose level when only 6 patients have been included at that dose. The design is recommended when the number of dose levels is limited (less than 6). To get around this criticism, one possible solution is to escalate the dose level after each patient has been treated as long as only mild toxicity has been observed. When the first moderate (not DLT) toxicity occurs, initial cohorts of 3 patients could then be treated before continuing to the next dose level. This is also known as “accelerated titration”.

Another point is that there is no a priori reason to use 3+3. If appropriately argued, different numbers may be envisaged.

The probability of selecting each dose level (operating characteristics under any n+m rule, such as 3+3) depending on the shape of the dose-DLT curve, can be simulated as needed.

3.2.3.2.2 Continual Reassessment Method (CRM) and other model based methods

The Continual Reassessment Method (CRM) is an attempt to improve the efficiency of the dose finding process. The dose levels to be tested are determined as explained above, for example using a Fibonacci sequence or even increments based on toxicity data from previous levels. The MTD is defined as the dose level associated with a predefined percent of DLTs (generally 20%. The general idea is to include patients at the best current estimate of the MTD. The design is divided into 2 stages, an initial escalation stage until the first DLT is observed, and then a second, model guided stage.

In the model free escalation stage, one could escalate the dose level either after each patient or every 3 patients have been treated as long as only mild toxicity has been observed. When the first moderate (not DLT) toxicity occurs, cohorts of 3 patients are treated before continuing to the next dose level.

After the first DLT, the dose for every new patient is recommended based on estimates of the probability of toxicity at each dose level. A model-based dose-toxicity curve is fit on ALL the previous data each time a new patient has been assessed. The estimation is obtained either using a Bayesian or a likelihood approach. This curve provides updated estimates of the probability of a DLT at each dose level. The accuracy of the estimate of the MTD improves as the sample size increases. The next dose level is selected as the dose whose estimate is closest to the target 20% (best current estimate of the MTD).

The main advantage of the CRM and other model-based approaches (also known as "adaptive dose-finding designs" in the literature) is to decrease the number of patients included at very low dose levels and to concentrate most of the dose levels actually used around the MTD. A confidence interval for the final estimate can also be calculated.

Details concerning the fitting of the CRM are given by O'Quigley et al (1999) and by Rosenberger and Haines (2002) who also consider the fitting of other designs for phase I trials.

For circumstances when one needs to also control for the risk of overdose (and toxicity), other model-based approaches are avaible, such as EWOC (Babb 1998), BLRM (Neuenschwander 2008). These methods function like the CRM, but the next dose is selected as the safest current estimate, i.e./ the dose adjacent to the tested one (no skip allowed) that has a probability to be "overdosing or unacceptable tox" < 25% (or specified risk)

For the circumstance when the event of interest is delayed in time, a time-dependent variant of the CRM (TITE-CRM) has been proposed by Cheung et al (2000). It allows including new patients at the best recommended dose even if the previous patient has not completed the evaluation period T and the model accounts for censoring.

3.3 Phase I trials for drugs for <u>which a increasing dose-toxicity and dose-activity relationship may not hold</u>

3.3.1 Objectives and endpoints

Modern drugs such as molecular targeted agents (MTAs), are often administered to patients in a chronic-fashion. Therefore patients are exposed to the drug for longer periods of time. Such agents require the assessment of the whole toxicity profile: on top of acute early toxicity, assessment of late toxicity profile (late and cumulative toxicity or chronic fashion-toxicities) is needed. Furthermore, adverse reactions at grades not regarded as dose limiting for agents intended to be used for a limited period of time, may not be

acceptable when the agent is to be given chronically (see Postel-Vinay et al 2011 and 2014) so that the definition of DLT may need to be adapted

For many such agents, the proposed mechanism(s) of action are not straightforward in that (1) the dose-efficacy curves are usually unknown and (2) dose-toxicity relationships are expected to be minimal, so that the paradigm underlying the designs discussed in 3.2 will not all apply. For illustration purposes, refer to the right hand panel of Figure 1. In such cases (biologic agents, vaccine, immunotherapies etc.), the toxicity is no longer a "surrogate" for the activity and it may then be important to include several patients at all dose levels in order to carry out translational research.

Translational research is essential since early phase studies should also aim to provide mechanistic proof of the concept underlying the drug action. Early studies must also document (through pharmaco-dynamic and pharmaco-kinetic endpoints) target modulation and drug effect in surrogate tissues; and provide information about potential predictive markers identifying a population of interest (often linked to the mode of action of the drug and putative target).

When using biological endpoints, one must pay attention to issues related to the complexity of cellular and signaling pathways, the use of tumor biopsy tissue or surrogate tissue, the optimal level of target expression/inhibition, the reliability of the assay for measuring the drug's effect, and the correlation with tumor response in animal models.

The trial may also serve to document any preliminary evidence of objective antitumor activity (clinical response) and to explore explore drug/drug (for combination studies) and/or food/drug interactions.

3.3.2 Dose escalation schemes

Although the classical designs are still often use for these agents, novel designs that incorporate two endpoints (one for safety and one for efficacy) have been proposed among which:

- TriCRM (Thall 1998) involves, combining two ordinal endpoints (one for safety and one for efficacy), into a new single ordinal response (2= activity and no toxicity, 1=no activity and no toxicity, 0= toxicity irrespective of activity), modelling the ordinal response and comparing doses on the basis of probability for answer "2" (activity and no toxicity)
- bCRM (Braun 2002) is built on models dealing with a bivariate endpoint (one for toxicity, and one for activity) and assessing its correlation. The model compares doses on the basis of their euclidian distance to a theoretical best point (e.g. prob of activity= 1 and prob of toxicity=0)
- EffTox (Thall 2004): bCRM like approach where comparison of doses is done on the basis of a tradeoff function that accounts for the correlation between activity and toxicity endpoints.

Whenever classical designs are used for early studies of MTAs so that only the toxicity during an initial and limited period of treatment is used for deciding on dose escalations, it is recommended to

- Confirm the safety of the selected RPIID by assessing the toxicity on a longer treatment period (eg 4 to 6 cycles or months)
- Consider including specific target-related toxicities and possibly some grade 2 toxicities known to be troublesome over longer periods of time (e.g. fatigue) in the definition of DLT

- ♦ Envisage an expansion cohort at the DLT to collect sufficient data on safety and sufficient material for translational research analysis and investigation of mechanism of action.

In general studies of this kind of agent will always will have attached translational research that allows to measure drug activity at the biological level.

3.4 Additional stopping rules and total study sample size

Contrary to other phase clinical trials, it is not possible to know in advance the number of patients required to complete a phase I study. It depends on:

(i) the number of escalation levels that are required to reach the MTD

(ii) the shape of the dose-toxicity curve

(iii) the design, and

(iv) the precision required at the end of the trial.

When using the standard method, a minimum of 6 patients are required at the final recommended level.

When using the CRM, different stopping rules are available. The 3 main ideas are:

- Stop after a given number (or proportion) of patients have been treated at the same dose level.
- Stop when the width of the 95% probability interval for the MTD becomes sufficiently narrow (adaptive design)
- Stop the trial when the probability to maintain the same dose level in the next ‘j’ patients exceeds some threshold value (80% for instance).

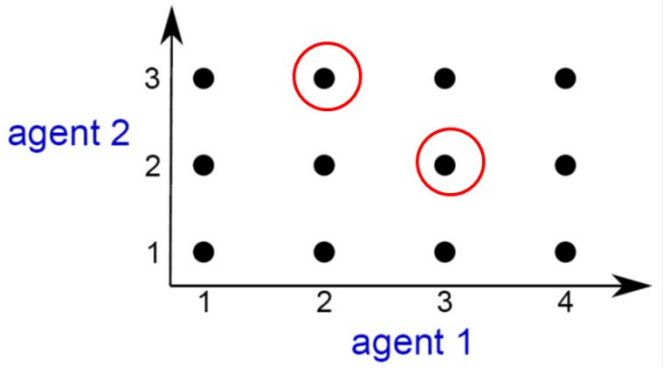
The decision of stopping a trial should be taken after discussion between the Statistician and Study Coordinator. It is important to carefully monitor the model based designs to detect circumstances when it would continue to indefinitely add more patients to the current dose level or to continually jump back and forth between two different dose levels.

3.5 Phase I trials of drug combinations

Whenever only one drug is incremented, the considerations reviewed earlier for single agent studies apply. However, every trial has to be designed based on prior knowledge of the preclinical and clinical pharmacology of the individual agents. Given the enormous number of potential combinations, in addition to a proper rationale, an appropriate hypothesis regarding the expected interaction at the pharmaco-dynamic and/or pharmacokinetic level should guide the protocol design. Hamberg et al (2010) discuss those aspects.

Specific dose-escalation designs are needed when the dose of both drugs is intended: complexity arises from the lack of obvious ordering of the dose levels. In the graph below multiple path of dose escalation are possible, and it is not obvious to declare if the pair of doses (level 2 of agent 1, level 3 of agent 2) represent a

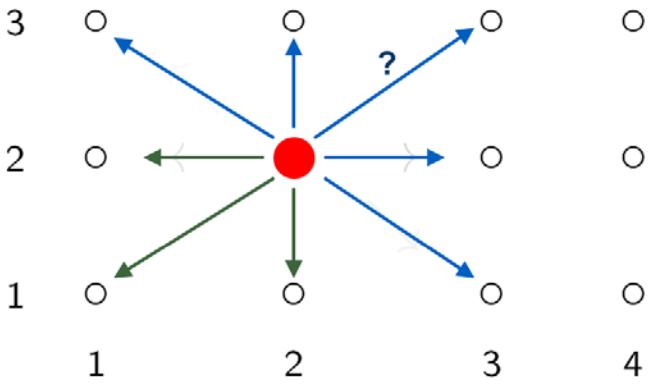
higher or lower dose level than (level 3 of agent 1, level 2 of agent 2).



The 3+3 design and other non-parametric algorithm-based designs (Ivanova et al 2003) make the assumption that the probability of toxicity of one agent is non-decreasing with dose when the dose of the other agent is fixed. These designs may not be appropriate when the number of dose levels to explore is large (the number of combination dose levels then becomes very large and the study risks requiring many patients if the RPIID is close to the maximum dose levels). Furthermore, if the agents are expected to have unknown dose-efficacy outcomes, designs that use dual endpoints capturing both activity and safety, such as those proposed by Zohar and Chevret (2007) may be more appropriate.

Mandrekar, Qin and Sargent (2010) and Mandrekar (2014) discuss designs appropriate for combination of agents and propose model-based designs appropriate for the purpose.

Model-based approach will recommend the next combination of doses, using the administered combination and the observations in order to decide which dose level will be allocated to the next cohort of patient. An initial escalation stage is predefined, and is used until the first DLT is observed. It may for example be predefined to escalate along the diagonal. After that time a model is used to estimate the probability of toxicity for all combinations of dose level, the next dose is the adjacent dose closest to the target with acceptable toxicity restrictions.



For a review of model-based approaches, you may also refer to Riviere et al (2014) who compare 6 recent dose escalation designs.

3.6 Studies involving irradiation

Clinical research in radiotherapy often follows a somewhat different approach.

When the aim is to investigate the delivery of higher doses using new therapeutic modalities (i.e. new algorithm or devices for delivering the irradiation) for tumors already known to be radiosensitive, the dose and the toxicity profile need to be closely investigated, but activity does not need to be demonstrated. Toxicity may however be deferred which will either prolong the observation period required for assessing the DLT, or method involving time-to-event endpoints (such as TITE-CRM) may be needed.

However, when a MTA or a radio-sensitizing agent is added to irradiation and when the dose escalation concerns that agent, the method described above will most often apply.

For a review of phase I trial designs in this setting see Pijls-Johannesma et al (2010)

3.7 Phase I/II trials

In some trials there is interest to look for a dose (Phase I part) and then within the same protocol proceed to test the selected RPII dose as an expansion phase of the phase I. The expansion being built according to a Phase II (usually non randomized) design. In such cases it is important to:

- Try to make sure the application of the Phase I selected dose can take place in the Phase II part without necessitating an amendment.
- Separate the design and the analysis of the two study parts clearly in the protocol.

For guidance on designing each part, refer to the present document and ST-001-WIN-02.

3.8 Softwares

The “Sample size tables for clinical studies” software developed by Machin can also be used for designing and monitoring phase I studies

Several R packages offer calculations for modern phase I designs (DFCrm).

- R Package dfcrm (http://cran.r-project.org/web/packages/dfcrm/) This package provides functions to run the CRM and TITE-CRM in phase I trials and calibration tools for trial planning purposes
- R package bCRM (http://cran.r-project.org/web/packages/bcrm/) for B CRM

EWOC can be found online on Cedar’s Sinal website (http://biostatistics.csmc.edu/ewoc/ewoc-s.php)

BLRM runs on WinBUGs.

Efftox, BCRm and more can be downloaded from MD Anderson cancer center website (Efftox, B CRM and much more) (https://biostatistics.mdanderson.org/SoftwareDownload/)

In 2003, when the EORTC had a phase I unit, the EORTC has also developed a software for the monitoring of trials using the classical CRM technique which is still located at V:\CRM.

4 References

- Babb, J., Rogatko, A. and Zacks, S. "Cancer Phase I clinical trials: Efficient dose escalation with overdose control," Statistics in Medicine 1998: 17:1103–1120.
- Bekele BN, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. Biometrics 2005; 61(2):343-54.
- Braun, T. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. Controlled Clinical Trials 2002; 23:240–256.
- Cheung K, Chappell R: Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics 2000; 56:1177-1182.
- Hamberg P, Ratain M, Lesaffre E, Verweij J. Dose-escalation models for combination phase I trials in oncology. Eur J Cancer 2010; 46: 2870-8.
- Ivanova A. A new dose-finding design for bivariate outcomes. Biometrics 2003; 59(4):1001-7.
- Mandrekar, S, Qin R and Sargent D. Model-based phase I designs incorporating toxicity and efficacy for single and dual agent drug combinations: Methods and challenges, Statistics in Medicine 2010; 29 1077-1083.
- Mandrekar R. Dose-Finding Trial Designs for Combination Therapies in Oncology. J Clin Oncol 2014; 32 (2): 65-7
- Neuenschwander B, Branson M, and Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. Statistics in Medicine 2008; 27:2420-2439.
- O'Quigley J, Shen Z, and Gamst A. Two-Sample Continual Reassessment Method. *Journal of* Biopharmaceutical Statistics 1999, 9: 17-44
- Pijls-Johannesma M, van Mastrigt G, Hahn SM et al. A systematic methodology review of phase I radiation dose escalation trials. Radiotherapy and Oncology 2010; 97: 135–141
- Postel-Vinay S, Gomez Roca C, Molife RL et al. Phase I Trials of Molecularly Targeted Agents: Should We Pay More Attention to Late Toxicities? J Clin Oncol 2011; 29:1728-1735
- Postel-Vinay S, Collette L, Paoletti X et al. Towards new methods for the determination of dose limiting toxicities and the assessment of the recommended dose for further studies of molecularly targeted agents – Dose-Limiting Toxicity and Toxicity Assessment Recommendation Group for Early Trials of Targeted therapies, an European Organisation for Research and Treatment of Cancer-led study. Eur J Cancer. 2014 ; 50(12):2040-9
- Riviere MK, Dubois F, Zohar S. Competing designs for drug combination in phase I dose-finding clinical trials.Statistics in Medicine 2014 Jan 27 [Epub ahead of print]
- Rosenberger WF, Haines LM. Compmeting designs foro phase I clinical trials. A review. Statitics in Medicine 2002; 21: 2757-2770.
- Thall P, Cook J. Dose-Finding Based on Efficacy-Toxicity Trade-Offs. Biometrics 2004; 60: 684–693.
- Thall P, Russell KE. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. Biometrics 1998;54(1):251–264

♦ Zohar S, Chevret S. Recent Developments in Adaptive Designs for Phase I/II Dose-Finding Studies. J Biopharm Stat. 2007;17(6):1071-83.

5 ASSOCIATED DOCUMENTS

None.

6 DOCUMENT HISTORY

Version N°	Brief description of change	Author	Effective date
1.00	Initial release, supersedes WP1102 version 1.5	Jan Bogaerts	25 Feb 2011
1.00	No change	Jan Bogaerts	25 Feb 2014
2	Revision to adapt to modern designs and molecular targeted agents	Laurence Collette	11 Dec 2015