Statement of searches: Pubmed was searched for the terms “early phase clinical trial design” and articles of relevance were individually reviewed.

Conflicts of interest: none declared
Early phase clinical trial designs - State of Play and Adapting for the Future

J.A. Harrington*, T.C. Hernandez-Guerrero*, B. Basu

* joint first authors

Cambridge University Hospitals NHS Foundation Trust, Addenbrooke’s Hospital, Department of Oncology, Box 193, Hills Road, Cambridge CB2 0QQ, UK

Corresponding author:

Dr Bristi Basu,

Honorary Consultant in Medical Oncology

Cambridge University Hospitals NHS Foundation Trust

Department of Oncology

Box 193

Addenbrooke’s Hospital

Hills Road

Cambridge

CB2 0QQ

T: 01223 769310, F: 01223 763120

E: bristi.basu@cruk.cam.ac.uk

Word count: 3699 (including abstract and sub-headings)

Number of tables: 3

Keywords: early-phase, Phase I trials, clinical trials, adaptive design, dose escalation methods, biomarker
Abstract
The process of anti-cancer drug development is complex, with high attrition rates. Factors that may optimise this process include highly accurate pre-clinical testing and using biomarkers for patient selection. However, the design of early phase clinical trials is likely to play a vital role for conducting robust clinical investigation of new targeted therapies and streamlining drug development. In this overview, we assess current concepts in Phase I clinical trials, highlighting issues and opportunities to improve their meaningfulness. The particular challenge of how to design combination trials is addressed, with focus on the potential of new adaptive and model-based designs.

Keywords (alphabetical order): adaptive design, biomarker, dose escalation methods, early-phase clinical trials, Phase I trials

Introduction
The development of anti-cancer drugs generally follows a conventional step-wise progression between phases of trials, each with different objectives but aiming to find signals that allow advancement to the next stage of clinical testing (Table 1). This process is widely recognised to be slow and inefficient, and ultimately less than 10% of new therapeutic agents are approved(1,2). Because of these high failure rates, there has been increased focus on strengthening the underlying pre-clinical work required to generate valid hypotheses(3) and using biomarkers to identify the most appropriate patients for treatment, particularly in the era of molecular targeted agents (MTA)(4). Alongside this, it is important to consider the design of early phase clinical trials, to ensure they incorporate rigorous stop-go signals and streamline drug development to enable new agents to fail promptly if they are not destined to be tolerable or active.

Increasingly pre-clinical work amasses data on pharmacodynamics (PD), pharmacokinetics (PK), toxicological profiles, dose or exposure/effect relationships and potential interactions. There are several examples of drug development where a PK/PD model has been used to both predict the therapeutic window and design the dosing schedule, with multiple trial designs
assessed *in-silico* before patients are treated(5). Window of opportunity studies, with short durations of drug administration are becoming increasingly popular, encouraging insight into novel therapeutic mechanisms of action at an early stage of a drug’s development.

Early phase clinical trials are generally defined as Phase I and non-randomised Phase 2 trials. This review focuses on Phase I trials, which include the first-in-human study of a new investigational medicinal product (IMP) as monotherapy, a combination of approved standard drugs, a combination of approved drugs and a new IMP, combination of IMP or combinations of new or approved drugs with radiation. Phase I trials aim to establish the optimal dose and schedule of a novel drug or combination of drugs, whilst determining the toxicity profile. The key endpoint is to determine the recommended phase II dose (RP2D) based on the determination of the maximum tolerated dose (MTD) of the IMP under investigation. Phase I study endpoints are summarised below:

- **Primary**
  - Identify MTD and RP2D
  - Identify dose limiting toxicities (DLTs)

- **Secondary**
  - PK
  - PD (molecular and clinical)
  - Target modulation
  - Efficacy

Traditionally phase I trials do not feature anti-tumour efficacy as a primary objective, and historically there has been a low probability of response (<10%) in early phase studies(1). Therefore patient motivation for entering such clinical trials must be considered, particularly in patients who have no standard treatment options available, and may have limited life expectancy. The informed consent process should explore the intensity of the required trial interventions, significant time commitment, potential for serious toxicity and low chance of achieving benefit for the individual concerned.
For many decades, these trials have assumed that a higher dose will be the most efficacious and that the probability of toxicity will increase with increasing dose. Whilst these assumptions may be valid with conventional cytotoxic agents, in the era of MTA and immunotherapies and with the increasing use of combinations, there is recognition that new designs should aim to identify the most active dose with the fewest adverse events rather than the MTD. In this review, we discuss early phase clinical trial designs, both as single agents and in combination, identifying the issues with conventional designs and the potential of alternatives.

Current concepts in Phase I trial design

The main principle guiding dose escalation in Phase I trials is to treat as many patients as possible within the therapeutic dose range, avoiding unnecessary exposure of patients to sub-therapeutic doses of an agent, while preserving safety and maintaining rapid accrual(6). During the escalation stage, patients are recruited into the trial sequentially in cohorts to receive a dose equal to or higher than the previous patient, with appropriate intervals between cohorts to perform safety reviews before opening the next higher dose cohort. Based upon the occurrence of severe toxicity, DLT, may be identified. DLT is defined prospectively as unacceptable adverse events (AEs), either due to severity (e.g. grade 3 or 4, determined by Common Terminology Criteria for Adverse Events (CTCAE)), or duration, which limits further dose escalation, and this is classically based on toxicity emerging in cycle 1 of treatment administration. More recent DLT definitions, however, may be more nuanced, such as Grade 3 gastrointestinal AEs despite adequate concomitant preventative medicine, or Grade 2 chronic and un-remitting toxicity. If the MTD is confirmed within a particular dose cohort, the RP2D may be defined based on pre-agreed criteria(7–9).

Starting dose

Historically the initial starting dose was selected based on rodent (mouse/rat) and non-rodent (dog/non-human primate) toxicology. Although myelosuppression and gastrointestinal toxicity in humans may be reflected using rodents, hepatic and renal toxicity is less reliably predicted,
making the use of a second species frequently necessary. For approximately 20% of new drugs, mouse data alone is insufficient to safely predict the human MTD(12). The dose (defined in mg/m² of body surface area) associated with 10% lethality in mice (MELD₁₀) can be predicted to be roughly equivalent to the human MTD(10), with the initial phase I trial dose 1/10 the MELD₁₀ or, if smaller, 1/3-1/6 the lowest dose that causes any toxicity (toxic dose low, TDL) in non-rodents(11). Another method to select the starting dose is use of the “no observed adverse effect level” (NOAEL). The starting dose for Phase 1 trials with a PK or PD endpoint is generally 1/50 the rat NOAEL(13). Allometric scaling is used to calculate equivalent surface area dose across species.

Escalation trial design

In general, there are two types of designs for the escalation process in early phase trials: rule- or model- based.

Rule-based

These rely on pre-specified rules that guide the escalation process throughout the entire trial development, and therefore cannot be changed. The design most commonly used is the 3+3 design. Patients are enrolled in groups (cohorts) of three. The escalation process assigns progressively increasing dose levels, pre-specified by the investigators. The dose escalations usually follow a modified Fibonacci sequence, enabling more conservative increments as higher, potentially more toxic dose cohorts are entered, although this is recognised to be inexact(14). The MTD is generally defined as the dose with a risk of ≥ 33% of patients experiencing a DLT, and the recommended phase 2 dose (RP2D) is immediately below that “toxic” dose. These rule-based designs are practical and easy to implement (“I see this, so will do this”), but escalation can be lengthy depending on the number of cohorts or drug combinations, and a large number of patients could end up receiving a dose below the final RP2D(15).
Variations of the standard 3+3 have been developed with the intent of speeding up the escalation process, such as the rule-based accelerated titration design. Here drug dose is doubled in single-patient cohorts until a grade 2 AE is experienced, at which point escalations revert to a standard 3+3 design (16). Intra-patient dose escalation may be used, whereby a patient treated at dose level 1 can be escalated to dose level 2 if dose level 2 is later well tolerated. This attempts to address the challenge of exposing cohorts of patients to sub-therapeutic doses, reducing the total number of patients required and enabling more rapid study progress. However, this may not be feasible when patients cannot stay on trial for a sufficient period, and introduces complexities in assessment of acute toxicities at a given dose level versus cumulative toxicity from a prior dose.

Whilst the traditional 3+3 approach worked well during the cytotoxic era, it is more problematic with the development of MTA and immunotherapies where the dose-efficacy curve can have a different shape. Higher doses may not correlate with greater efficacy (17), severe toxicities may appear later after cycle 1 and be more chronic in nature, particularly since MTA and immunotherapy administrations may continue for many treatment cycles until disease progression (18, 19). Moderate (grade 2) toxicities that recur, or are prolonged must be considered since they could significantly impair quality of life (20). An even greater challenge is presented when combining new therapies with radiation therapy, as radiation toxicities can appear very late after dosing, leading to a longer follow-up and delay in accrual, thus increasing costs and complicating the analysis of results. In these trials, it is critical that the study population is carefully defined including the radiotherapy volumes, dosing and planning techniques and that normal tissue toxicities are assessed (21, 22).

Model-based

These designs use the information accrued during the trial to guide escalation (23). The dose-toxicity relationship is characterised by one or more parameters, with re-assessment of this during the trial, allowing decision making about the best dose level to be assessed. Many of these follow a Bayesian method which generates probabilities of DLTs for the compound,
based on prior information obtained from pre-clinical or clinical data(24,25). The best known is the continual reassessment method (CRM), where the model requires the dose of the drug and a single model parameter(26). Based on the prior distribution for the parameter and previous DLT responses, the drug-toxicity relationship can be updated and the next dose selected as that predicted to be closest to the toxicity criteria. A benefit of Bayesian methods is that the incorporation of prior information aids the dose-escalation decision-making process, meaning fewer patients receive a dose below the RP2D, making the process more efficient. Moreover, adaptive or model based designs can be ideal for addressing many questions at once. Using all the information gathered during the trial, the design can help identify the appropriate patient population, dose and regimen, and therapeutic combinations.

**Current practice**

Whilst there are a range of rule based designs, a review of six of these (including the 3+3) found that the cumulative group up-and-down design was the best performing of these, likely to give similar results to a model based design(27). Both rule- and model-based designs can be modified, adjusting the cohort size or testing intermediate doses, to enable flexibility if unexpected DLTs occur. Evaluating PK and PD effects on target and pathway can be used to inform dose and schedule(28,29). Whilst these changes can be pre-specified particularly in model-based designs, they may require protocol amendment (with accompanying delays) for rule-based designs. Modifications to the CRM can adapt the starting dose, cohort size and the number of dose levels that can be escalated between cohorts(30).

The vast majority of early phase trials follow a rule-based design. In a recent review, 93% of dose-finding trials utilised a rule-based design, although an increase in model-based designs was seen between 2009-2014(31). Indeed most trials testing MTA or immunotherapies still use the standard rule-based design assessing toxicity for MTD determination, in spite of their drawbacks. Table 2 summarises some of the advantages and disadvantages of different designs.

**Newer concepts in trial design**
PD guided dosing

Given the limitations of using DLT within first cycle of treatment as a guide for determining the RP2D, if PK and PD surrogate information on target engagement is available in a dynamic manner, an ideal model would recommend RP2D based not only on toxicity, but also efficacy, using evidence of biological activity (32,33). This should minimise toxicity occurring from off-target effects at higher concentrations of the drug. The concept of biologically active dose was initially designed to indicate quantitatively the biological effect of any radiotherapy treatment, taking account of changes in dose-per-fraction or dose rate, total dose and overall time(34). A similar concept is used for MTA, known as the Optimal Biological Dose (OBD), which refers to the dose associated to the most desirable effect on a specific biomarker such as inhibition of a key target in tumour or surrogate tissue (35). However, identifying the OBD can be complex due to factors such as effects on normal tissue and the variability in individual patient exposure to drug (36). Whilst its use has gained popularity in clinical development, it still remains as a secondary objective for most early phase trials, behind toxicity assessment(37).

Biomarker studies

In recognition that biological activity depends on the characteristics of each tumour type, which can be variable due to heterogeneity(38,39), PD studies are increasingly being incorporated within phase I trials of novel IMPs, to confirm and quantify target engagement, biochemical pathway modulation, and biological effects(40). This process can also identify potential markers of response, that could allow selection of patients that may benefit the most from a specific MTA(3,41). However, a biomarker needs to be robustly predictive to restrict eligibility solely to patients with that biomarker. For example, V600E BRAF mutations are highly predictive of response to BRAF and MEK inhibition in metastatic melanoma(42) but not all BRAF-mutated cancers are sensitive to BRAF and MEK inhibition(43), highlighting the context dependence of potential biomarkers.
Expansion cohorts

Given the small number of patients usually assessed in Phase I trials, concerns have been raised about the validity of the dose chosen as RP2D. Therefore studies may open pre-planned expansion cohorts where patients are enrolled at the RP2D, often with strict eligibility criteria designed to focus on either particular cancer subtypes or molecular tumour characteristics. Additional PK, PD, efficacy and tolerability information can then further inform subsequent Phase II trials, and identify particular patient populations which may benefit. A review of phase 1 trials run between 2006 and 2011 found that 24% included an expansion cohort. Amongst these, the RP2D was modified in 13% with new toxicities seen in 54% (44). However, it was unusual for these cohorts to identify a tumour response that had not been seen in the initial Phase I trial. Whilst use of expansion cohorts can provide valuable additional information on the IMP, they should not be considered a replacement for Phase II studies with valid, statistically determined endpoints (45).

Improved patient selection

With a limited number of patients potentially fit enough to enter clinical trials, careful selection of patients with adequate reserve may improve their ability to tolerate unknown toxicities of novel agents, thus reducing the potential for early drop-out within the conventional DLT period (cycle 1). This is incorporated within standard phase I inclusion/exclusion criteria for Phase I trial entry (see Table 3). For example, a multi-variate analysis of factors associated with survival in Phase I clinical trial patients showed that low albumin, raised lactate dehydrogenase and greater than two sites of metastatic disease were independently predictive of poor overall survival (46). A nomogram predicting risk of serious drug related toxicity in cycle one has value, for example identifying that some commonly used eligibility criteria are unlikely to impact on patient safety, such as baseline haemoglobin, whilst others such as albumin are better predictors (47). However, the potential pool of patients suitable for Phase I trials is already highly selected compared to the typical cancer patient population and further selection risks underestimating real-world toxicity. Whilst improved patient selection
and expansion cohorts are valuable additions to Phase I clinical trials, use of alternative trial design may offer an important role in improving patient safety (48).

**Alternative adaptive trial designs**

The statistical community are actively developing alternatives to the standard trial designs, although the majority remain untested in trials. Adaptive designs include the modified toxicity probability interval (mTPI) method. Here, no priors are required and Bayesian models are used to describe the observed toxicity data, with a set of decision rules based on toxicity posterior intervals (49). It is simple to implement and studies have shown good operating characteristics when compared to 3+3 including when sample size is matched, although it is only applicable to trials with a binary toxicity endpoint. It has been used in Phase I oncology trials, with a first stage standard 3+3 dose escalation to obtain toxicity data, followed by the use of the mTPI design, for example in a Phase I trial of a pan-AKT inhibitor MK-2206 (50).

To address concerns about how to incorporate delayed toxicity assessments, a range of adaptive designs have been proposed. These can include time-to-event (TITE) outcomes, e.g. TITE-CRM (51,52), and late onset toxicity responses (53,54). It is also possible to incorporate convert graded toxicity information into numeric scores (49,55) or consider cumulative toxicity over a number of treatment cycles (56). Trials where intra-patient dose-escalation or schedule changes are permitted (7), together with other designs where the likelihood that the particular AE is related to the drug can be integrated (57). One complex design combines escalation with overdose control, and assesses multiple toxicities per patient rather than a binary indicator of DLT together with TITE toxicity data (58), with simulations indicating this could improve the accuracy of MTD determination and shorten trial duration. Heterogeneity between patients, which may be influenced by many factors such as enzyme or receptor levels (for example HER-2 in breast cancer) can lead to varying efficacy and toxicity between patients treated at the same dose level. Therefore, further modifications to incorporate predictive patient biomarkers to specific adverse events and response as covariates may permit estimation of a “personalised” MTD in the future (59).
Efficacy in Phase 2 trials is usually defined by the change in tumour size on imaging ("response"), usually performed after toxicity assessments at the end of cycle 1. Hence, toxicity endpoints alone are still used to define RP2Ds in the majority of phase 1 trials(60). Designs developed to assess both toxicity and efficacy include one which initially evaluates the RP2D (achieved using either 3+3, accelerated titration or CRM designs) but then in the second step randomises patients to three dose levels – the RP2D, the dose level immediately below and that immediately above that level. The final RP2D is then determined using both toxicity and efficacy endpoints(61). Another models three probabilities: the probability of no response (no efficacy and no toxicity), the probability of success (efficacy and no toxicity) and the probability of toxicity (toxicity present regardless of efficacy) and uses prior information and evolving data for toxicity and efficacy to recommend the next dose (62).

Combination clinical trial design
The design of combination clinical trials is a particular challenge and requires a strong scientific rationale, with pre-clinical data, consideration of PK/PD and overlapping toxicities(63). Factors influencing toxicity in combinations include drug-drug interactions, for example altering drug clearance, thereby affecting the exposure or maximal concentration of a drug when combined. Overlapping toxicities may arise because both drugs have similar toxicity profiles and the combination increases the observed rate of a particular AE significantly. Combination Phase I trials may consider a single RP2D combination or multiple RP2D combinations, amalgamating a “surface of interactions” on the dose curves of different drugs to be evaluated(64). A pragmatic decision may be made that one dose/schedule is preferred but more than one RP2D combination with similar tolerability and efficacy could then be compared in a Phase II trial.

For the vast majority of combination trials, the dose of one drug is fixed (usually the drug about which most is known) with escalation of the novel agent, aiming to reach its single agent RP2D, using the 3+3 design and therefore only assessing a limited number of doses. However, this approach is complicated as the fixed dose agent may already cause toxicity (particularly in
heavily pre-treated oncology patients) and therefore DLT can prevent dose escalation of the novel agent, before target inhibition. Alternative rule- and model- based designs for combination Phase I studies have been proposed to take account of problematic overlapping toxicities(65). A modified up-and-down sequential design, with simultaneous escalation and de-escalation of IMPs to identify a MTD contour was used in a Phase I study of neratinib and temsirolimus, exploring twelve dose combinations and identifying two separate MTD combinations(66). Several model-based designs involve a start-up stage to inform the prior, often via a simple rule-based escalation, followed by the second stage using the model(67–70). However, a comparison of the performance of different adaptive models during dose escalation showed that designs where a single agent is escalated first whilst the other is fixed are inefficient and identify fewer RP2D combinations(71). An alternative Bayesian design allows dose escalation of both drugs with continuous dose levels, alternating the use of single agent conditional on the dose level of the other agent(72). More complex models have also been proposed(73–75), including one where surrogate measures of efficacy can be used until confirmatory efficacy data is obtained from imaging, allowing improved dose selection and shorter trial duration(76) and another where both efficacy and toxicity are modelled as TITE outcomes, using partial information to assign new patients to a dose level rather than requiring the final outcome from the previous cohort(77).

Conclusions
Phase I trials are critical for evaluation of new investigational anti-cancer therapies and should be conducted safely so the minimum number of patients are exposed to risk of serious toxicities; efficiently to minimise the number of patients exposed to potentially sub-therapeutic doses of drug, escalate tested doses rapidly in the absence of toxicity and more slowly when toxicity emerges; and dependably so there is confidence that the MTD holds across the population(78). They are generally conducted in multiple tumour types, and in patients who have no further standard of care options available, unless the studies include an approved drug, in which case they may be appropriate earlier in the treatment pathway. Expansion cohorts may concentrate on defined tumour types or patients positive for proposed
biomarkers, and biomarker studies can allow concepts such as the OBD of new agents to be explored. Whilst the majority of trials still follow standard rule-based designs, newer adaptive designs incorporating information beyond the traditional DLT period, and integrating pre-clinical data, information from other trials and emerging safety data are increasingly being explored with encouragement from regulatory agencies(79,80), despite the logistical challenges associated with this approach.
Table 1: Phases of clinical trials during drug development

<table>
<thead>
<tr>
<th>Phase 0</th>
<th>Very small exploratory trials carried out to determine the preliminary PK and PD characteristics of the new compound after administering limited doses, as a preliminary investigation before taking it to further evaluation (not routinely undertaken).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Small trials, non-randomised. Primary objective to assess safety and tolerability and achieving a recommended-phase-2-dose. They evaluate PK and PD biomarkers obtained with the different dose-schedules-combinations.</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Larger trials. Usually non-randomised. Main objective to assess anti-tumour activity in a specific setting (usually measure response rates as primary endpoint).</td>
</tr>
<tr>
<td>Phase 3</td>
<td>Very large trials. Mostly randomised. Primary objective is usually determining efficacy of the drug as compared to placebo/standard of care (if any). The primary endpoints are generally overall survival/progression free survival.</td>
</tr>
<tr>
<td>Phase 4</td>
<td>“Real life patients”. Post-marketing trials, testing long term safety in patient population.</td>
</tr>
</tbody>
</table>
Table 2: Advantages and disadvantages of rule-based and model-based designs

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Rule-based designs</th>
<th>Model-based designs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplicity of implementation</td>
<td>Simple to implement.</td>
<td>More complicated, require constant input from a statistician who analyses the data.</td>
</tr>
<tr>
<td>Drug exposure for patients.</td>
<td>Significant number of patients treated at “sub-therapeutic” doses.</td>
<td>Smaller number of patients treated at sub-therapeutic doses.</td>
</tr>
<tr>
<td>Costs</td>
<td>Not particularly expensive.</td>
<td>Can be more expensive because of development requirements.</td>
</tr>
<tr>
<td>Time for implementation</td>
<td>Can take long time to complete.</td>
<td>Allow a more rapid dose escalation process.</td>
</tr>
<tr>
<td>Investigators’ knowledge</td>
<td>Largely known for researchers.</td>
<td>Less known so increased education required.</td>
</tr>
<tr>
<td>Assessment of toxicities in order to determine an MTD</td>
<td>Binary assessment / Presence or absence of DLTs as unique guide for MTD.</td>
<td>Allow the use of PK, and PD combined with toxicity data to determine a MTD, therefore more accurate.</td>
</tr>
<tr>
<td>Objective</td>
<td>Allow a single MTD to be “discovered” and potentially taken to further evaluation.</td>
<td>When used in combinations, can lead to several MTDs or RP2D, which can potentially be compared in Phase II trials or that with best PK/PD characteristics taken forward.</td>
</tr>
</tbody>
</table>

*Dose Limiting Toxicity; **Maximum Tolerated Dose*
Table 3: Phase I patient population eligibility criteria

<table>
<thead>
<tr>
<th>Conventional eligibility criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Advanced solid tumours, unresponsive to standard therapies or for which there is no effective treatment</td>
<td></td>
</tr>
<tr>
<td>• Performance status ECOG 0 or 1</td>
<td></td>
</tr>
<tr>
<td>• Adequate organ function (FBC, creatinine, AST/ALT, bilirubin)</td>
<td></td>
</tr>
<tr>
<td>• Specified prior therapy permitted</td>
<td></td>
</tr>
<tr>
<td>• Specified time interval between prior treatment and initiation of trial therapy</td>
<td></td>
</tr>
<tr>
<td>• No serious uncontrolled medical or psychiatric disorder or active infection</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial specific eligibility criteria can include:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Restricted patient populations – scientific rationale</td>
<td></td>
</tr>
<tr>
<td>• Specific organ function restriction e.g. QTc, LVEF, uncontrolled hypertension or proteinuria for anti-angiogenics</td>
<td></td>
</tr>
<tr>
<td>• Prohibited medications if significant risk of interactions</td>
<td></td>
</tr>
</tbody>
</table>


52. Huang B, Kuan PF. Time-to-event continual reassessment method incorporating treatment cycle information with application to an oncology phase I trial. Biometrical Journal. 2014;


http://dx.doi.org/10.1038/nrclinonc.2016.96%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/27377132

http://ascopubs.org/doi/10.1200/JCO.2012.47.2787


