Trial Designs

Which design is “Fit for Purpose”

Richard F. Schlenk, M.D.

National Center of Tumor Diseases (NCT)
German Cancer Research Center Heidelberg
Selected Topics in an Overall Exciting Field

- Phase I rules or models?
- Seamless designs more answers?
- Historical controls reduce sample size?
The primary aim of Phase I Studies in Oncology is to understand the toxicity properties of the new anti-cancer drug/strategy. The drug efficacy is a secondary endpoint especially in cross entity studies for signal detection. Treatment-related dose limiting toxicities (DLTs) define the endpoint for the individual patient and the whole study. DLTs are monitored traditionally for the first cycle, but longer time periods may be more appropriate. Based on the assumed correlation between dosage and induced side-effects Phase-I studies are traditionally dose-escalation studies. The number of patients with DLTs in each dose-level is used to determine the Maximally Tolerated Dose (MTD) defining subsequently the recommended dose for phase-II studies (e.g. cytotoxic drugs). With molecularly target therapies MTD may not be reached and other read-out are necessary to determine the recommended dose for follow-up studies.

Ananthakrishnan R et al. Contemporary Clinical Trials Communications 5 (2017) 34e48
# Phase-I Designs - Rule-based

<table>
<thead>
<tr>
<th>Design</th>
<th>Assignment rule</th>
<th>Ways to escalate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3 + 3$</td>
<td>If 0 out of 3 enrolled patients have a DLT, then escalate to the next dose level and enroll 3 more; if 1 out of 3 patients has a DLT, then add 3 more patients at the same dose level; if 2 or more patients out of 3 or 6 patients experience a DLT, then stop the trial. The MTD is one dose level below.</td>
<td>$0/3 = 0%$ or $1/6 = 16.7%$ i.e. can escalate if we observe 0 DLTs out of 3 patients, or 1 DLT out of 6 patients</td>
</tr>
<tr>
<td>$2 + 4$</td>
<td>If 0 out of 2 enrolled patients have a DLT, then escalate to the next dose level and enroll 2 more; if 1 out of 2 patients has a DLT, then add 4 more patients at the same dose level; if 2 or more patients out of 2 or 6 patients experience a DLT, then stop the trial. The MTD is one dose level below.</td>
<td>$0/2 = 0%$ or $1/6 = 16.7%$ i.e. can escalate if we observe 0 DLTs out of 2 patients, or 1 DLT out of 6 patients</td>
</tr>
<tr>
<td>$4 + 4$</td>
<td>If 0 out of 4 enrolled patients have a DLT, then escalate to the next dose level and enroll 4 more; if 1 or 2 out of 4 patients have a DLT, then add 4 more patients at the same dose level; if 3 or more patients out of 4 or 8 experience a DLT, then stop the trial. The MTD is one dose level below.</td>
<td>$0/4 = 0%$ or $1/8 = 12.5%$ or $2/8 = 25%$ i.e. can escalate if we observe 0 DLTs out of 4 patients, or 1 DLT out of 8 patients, or 2 DLTs out of 8 patients</td>
</tr>
<tr>
<td>$5 + 5$</td>
<td>If 0 out of 5 enrolled patients have a DLT, then escalate to the next dose level and enroll 5 more; if 1 or 2 out of 5 patients have a DLT, then add 5 more patients at the same dose level; if 3 or more patients out of 5 or 10 experience a DLT, then stop the trial. The MTD is one dose level below.</td>
<td>$0/5 = 0%$ or $1/10 = 10%$ or $2/10 = 20%$ i.e. can escalate if we observe 0 DLTs out of 5 patients, or 1 DLT out of 10 patients, or 2 DLTs out of 10 patients</td>
</tr>
<tr>
<td>$3 + 3+3$</td>
<td>If 0 out of 3 enrolled patients have a DLT, then escalate to the next dose level and enroll 3 more; if 1 out of 3 patients has a DLT, then add 3 more patients at the same dose level; if 2 out of 6 patients have a DLT then add 3 more patients at the same dose level; if 2 or more patients out of 3 patients experience a DLT or 3 or more out of 6 or 9 patients experience a DLT, then stop the trial. The MTD is one dose level below.</td>
<td>$0/3 = 0%$ or $1/6 = 16.7%$ or $2/9 = 22.2%$ i.e. can escalate if we observe 0 DLTs out of 3 patients, or 1 DLT out of 6 patients, or 2 DLTs out of 9 patients</td>
</tr>
</tbody>
</table>

**Simple Accelerated Titration Design**

Successively assign a single patient at each dose level until the patient has a DLT. Then switch to the $3 + 3$ design (i.e. add 2 more patients to the dose level at which a DLT is first seen and then follow the rules of the $3 + 3$ design).
Phase-I Designs - Model-based

- **Modified Toxicity Probability Interval (mTPI)**
  - Based on a Bayesian dose finding design that uses posterior probability to guide dose selection -> Unit probability mass (under - proper – overdosing)

- **Toxicity Equivalence Range (TEQR)**
  - Frequentist version of the mTPI.

- **Bayesian Optimal Interval Design (BOIN)**
  - In contrast allows for flexible interval based on dose level and number of patients in this dose level

- **Continuous Reassessment Method (CRM)**
  - Uses the DLT information from all previous patients to determine the dose level to which the next patient is assigned based on the dose-toxicity model backbone

- **Escalation with Overdose Control (EWOC)**
  - Based on an adaptive Bayesian dose finding design with overdose control as unique feature
Operational Characteristics according to dose-tox-model

**logistic dose-toxicity:** 
\[ \log_e(\text{DLT rate}/(1-\text{DLT rate})) = -5.96641 + 0.013713 \times \text{dose}. \]

**linear dose-toxicity:** 
\[ \text{DLT rate} = \min(-0.071197 + 0.000811966 \times \text{dose}, 1). \]
Trial design

- e.g. dose level 3 is MTD:

![Graphs showing experimentation and recommendation with dose levels 1 to 5 and percentages for methods 3+3 and CRM.]

Courtesy: A. Kopp-Schneider (Inform² biostatistician, DKFZ Heidelberg)
Helpful Operational Characteristics for phase-I trials

- % of times that dose level 3 is selected as the MTD
- % of times that doses below the MTD (dose levels 1 and 2) are selected as the MTD
- % of times that doses above the MTD (dose levels 4 and above) are selected as the MTD
- Average number of dose levels examined
- Std of dose levels examined
- Max dose levels examined
- Median dose levels examined
- Average number of patients per trial
- Median number of patients per trial
- Median number of DLTs per trial
- Average sample size at MTD
- Average % of pts dosed at MTD
- Average % of pts under- dosed
- Average % of pts over- dosed
8- bis 12-jährige Entwicklungsphase

Entwicklungskosten: 300 bis 800 Mill. Euro

Zulassung

Phase III
(2 bis 4 Jahre)

Wirksamkeit
Verträglichkeit

Phase II
(1 bis 1,5 Jahre)

Verträglichkeit
Immunogenität

Phase I
(1 bis 1,5 Jahre)

Dosis
Verträglichkeit

Erforschung und präklinische Entwicklung
(2 bis 5 Jahre)

Identifizierung und Herstellung
der Antigene
Pharmakologie

Frühe Entwicklungsphase

Fortgeschrittene Entwicklungsphase

http://stage.spmsd.de/impfstoffe/sicher-geimpft/
Example: 2011 Start of the Phase-I Study NCT01295827
- Study of Pembrolizumab (MK-3475) in Participants With Progressive Locally Advanced or Metastatic Carcinoma, Melanoma, or Non-small Cell Lung Carcinoma (P07990/MK-3475-001/KEYNOTE-001) (KEYNOTE-001)
- Accelerated Approval 04SEP2014 of Pembrolizumab in Melanoma
- Accelerated Approval 02OCT2015 of Pembrolizumab in NSCLC

All clinical trials should have clearly stated objectives, with a design and statistical analysis plan capable of achieving those goals.

The type of attention to patient protections afforded by conventional, phased trial designs can be incorporated into this approach through more careful selection of the drugs to be studied in this fashion, greater attention to the statistical rationale and analysis plan for additional cohorts, establishment of external oversight committees, and more frequent, real-time communication among sponsors, investigators, IRBs, regulators, and patients.

PHASE III RANDOMIZED TRIAL OF VOLASERTIB PLUS LOW-DOSE CYTARABINE (LDAC) VERSUS PLACEBO PLUS LDAC IN PATIENTS AGED >65 YEARS WITH PREVIOUSLY UNTREATED AML, INELIGIBLE FOR INTENSIVE THERAPY (EHA 2016)

Pts were randomized 2:1 (stratified by ECOG [0/1 vs 2] and type of AML [de novo vs secondary]) to receive LDAC (20 mg s.c. BID Days 1–10 Q4W) and either V (350 mg; 1-hr iv infusion Days 1 and 15 Q4W) or placebo (P). The primary analysis was performed after completion of recruitment (Nov 2014) and focused on efficacy in pts randomized ≥5 months before clinical cut-off. Objective response (OR; complete response [CR] + CR with incomplete hematological recovery [CRi]; blinded central review) was the primary endpoint, and OS was the key secondary endpoint. An additional OS analysis was performed in all randomized pts (Nov 2015).
PHASE III RANDOMIZED TRIAL OF VOLASERTIB PLUS LOW-DOSE CYTARABINE (LDAC) VERSUS PLACEBO PLUS LDAC IN PATIENTS AGED >65 YEARS WITH PREVIOUSLY UNTREATED AML, INELIGIBLE FOR INTENSIVE THERAPY (EHA 2016)

666 pts were treated (444 V+LDAC; 222 P+LDAC). The primary analysis included 371 pts (246 V+LDAC; 125 P+LDAC); pt characteristics were balanced: median age, 75/75 yrs; secondary AML, 47%/49%; adverse genetics, 32%/32%, respectively. The percentage of pts with OR was not statistically significantly higher with V+LDAC vs P+LDAC (25.2% vs 16.8%; Odds ratio 1.66 [95% CI 0.95–2.89; p=0.071]), and thus the primary endpoint was not met. A negative OS trend was seen for V+LDAC vs P+LDAC (median OS: 4.8 vs 6.5 mos; HR 1.26 [95% CI 0.95–1.67; p=0.113]). Adverse event severity increased with V+LDAC vs P+LDAC, with fatal infection frequency of 16.6% vs 5.1%, considered to be the main reason for the negative OS trend. Consequently, the study was unblinded, and investigators/pts could stop/continue treatment based on individual benefit-risk evaluations.
PHASE III RANDOMIZED TRIAL OF VOLASERTIB PLUS LOW-DOSE CYTARABINE (LDAC) VERSUS PLACEBO PLUS LDAC IN PATIENTS AGED >65 YEARS WITH PREVIOUSLY UNTREATED AML, INELIGIBLE FOR INTENSIVE THERAPY (EHA 2016)

Estimated fatal infection frequencies

25% recruitment
- 18/111 (16.2%) 3/56 (5.4%)

50% recruitment
- 36/222 (16.2%) 6/111 (5.4%)

75% recruitment
- 54/333 (16.4%) 9/166 (5.4%)

100% recruitment
- 73/444 (16.4%) 12/222 (5.4%)
PHASE III RANDOMIZED TRIAL OF VOLASERTIB PLUS LOW-DOSE CYTARABINE (LDAC) VERSUS PLACEBO PLUS LDAC IN PATIENTS AGED >65 YEARS WITH PREVIOUSLY UNTREATED AML, INELIGIBLE FOR INTENSIVE THERAPY (EHA 2016)

Estimated fatal infection frequencies

25% recruitment
- 18/111 (16.2%) 3/56 (5.4%) 0.05

50% recruitment
- 36/222 (16.2%) 6/111 (5.4%) 0.005

75% recruitment
- 54/333 (16.4%) 9/166 (5.4%) 0.0005

100% recruitment
- 73/444 (16.4%) 12/222 (5.4%) 0.00004
Population – Selection I

- The **target population** describes a subgroup within the general population defined by a specific condition (e.g. disease, symptoms)
- **Inclusion criteria** define this target population of a specific trial
- **Internal validity** is based on the design of a clinical study (e.g. equality in structure and observation, randomization, blinding,…)
- **External validity** describes whether results of a clinical trial are generalizable to persons other than the population in the original study
- In- and exclusion criteria are the gateway into a clinical trial
- Deviations from in- and exclusion criteria are probably **major** or even **critical** protocol deviations
- Thus strict adherence to in- and exclusion criteria is mandatory
In- and Exclusion Criteria

- Define the trial population
- Represent selection criteria which are effective in between the target population and the clinical trial population
- Are key elements to get an idea of trial’s external validity
- Should be broad enough to allow as much as possible a representation of the target population in the study population
- Should be narrow enough to guarantee patients safety
Population – Selection

General Population

Trial Population

Analyzed Population

Inclusion Exclusion Criteria

Selection during trial conduct

Generalizability?

External validity
A seamless Phase IIB/III adaptive outcome trial: Design rationale and implementation challenges
ClinicalTrials.gov Identifier: NCT00543543

Vaccine characteristics
- **4vHPV**: HPV 6/11/16/18  
  70% of cervical cancers
- **9vHPA**: HPV 6/11/16/18 + 31/33/45/52/58  
  90% of cervical cancers

A 9-Valent HPV Vaccine against Infection and Intraepithelial Neoplasia in Women
ClinicalTrials.gov Identifier: NCT00543543

AMLSG 12-09: Trial Design – Phase-II

1st Induction  2nd Induction  Consolidation  Maintenance

Fast molecular Screening 48h

*Randomization, #Allocation according to initial randomization
**Inclusion:** mainly AML with MDS-related changes, AML not otherwise classified

**Exclusion:** AML with t(15;17), inv(16), t(8;21), NPM1mut, FLT3mut

All patients eligible for intensive chemotherapy, no upper age limit

*Randomisation, #Allocation according to initial randomisation
Taking Forward Successful Arms from Randomized Phase-II into Phase-III

Patients randomized into treatment arms A and D in the first four arm phase of the AMLSG 12-09 study are included into the evaluation of the second phase by left truncation.

Keiding N. Statistics in Medicine 2006;25:2343-2364
Importantly, it is crucial to include an adequate number of patients into a clinical trial.

Only if sufficient patients are included, there is a realistic chance that the trial is successful.

Each additional patient increases the recruitment times and costs of the study.

The availability of new therapies for affected patients gets postponed.

Recruitment of the required sample size is especially challenging in the field of rare diseases.

An attractive option: inclusion of historical data of previous (pilot) trials.
Non-Bayesian approaches

1. **Separate analysis** 153 patients per group required
   - we ignore the historical data
   - viewed as "standard analysis"
   - perform a one-sided two-proportion z-test

2. **Pooling** 115 patients per group required
   - we just pool the historical and the current data
   - equal randomization in the current trial
   - pool the historical controls with the current controls as if they had been observations in the current trial
   - perform a one-sided two-proportion z-test

3. **Test-then-pool** 115 patients per group required
   - make a choice between the "separate" and the "pooling" options
   - first perform a test of $H_0 : \pi_h = \pi_C$ vs. $H_1 : \pi_h \neq \pi_C$
     (z-test, $\alpha = 0.05$)
   - $H_0$ is rejected $\rightarrow$ perform a separate analysis
   - $H_0$ not rejected $\rightarrow$ perform a pooled analysis
Bayesian approaches with dynamic borrowing I: Hierarchical model (Meta-Analytic approach)

- Place a distribution across the true response rates in the different studies.
- $\pi_C$ true control rate in the current study
- $\pi_1, ..., \pi_H$ true control rates in the $H$ historical studies ($H$ maybe 1)
- Define $\gamma_0, \gamma_1, ..., \gamma_H$ to be the logits of the true control rates.
- Further assume
  \[ \gamma_0, ..., \gamma_H \sim \mathcal{N}(\mu, \tau^2) \]
- $\mu$ and $\tau$ present the between-study mean and standard deviation
- Add a second layer to the model (hierarchical structure)
  \[ \mu \sim \mathcal{N}(\mu_0, \tau_0^2) \text{ and } \tau^2 \sim IGamma(\alpha, \beta) \]
- Hierarchical model approach is very sensible to the choice of the prior for the between-study variance.
- Thus not well suited when only one or few historical studies should be integrated.
Wissenskontrolle TED-I
Welche Antwort ist korrekt?

• Der primäre Endpunkt in Phase-I Studie ist in der Regel Wirksamkeit
• Die “Operating Characteristics” verschiedener Phase-I Studiendesigns spielen für letztendliche Auswahl keine Rolle, da 3+3 immer am besten “performed”
• Mit dem Modell-basierte „Continuous Reassessment Method (CRM)“ Design erreicht man in der Regel mit einer höheren Wahrscheinlichkeit die korrekte Definition der Maximal-Tolerierte Dosis
• Die Dosis-limitierende Toxizität (DLT) und die Maximal-Tolerierte Dosis (MTD) sind immer in ein und derselben Dosisstufe
Die Entwicklung von neuen Substanzen mit einer „Breakthrough Designation“ also einer bisher nicht beobachteten Wirksamkeit ist mit Seamless-Studiendesigns effektiv und rasch bis zur Zulassung möglich.

Die Überwachung der Sicherheit für die teilnehmenden Patienten ist in einem Seamless-Design wie im Standardentwicklungsprogramm mit Phase I bis –III in gleicher Weise etabliert.

Die Verallgemeinerbarkeit der Studienergebnisse aus Studien mit Seamless-designs kann im Vergleich zum Standardentwicklungsprogramm mit Phase I bis –III deutlich eingeschränkt sein.

Der nahtlose Übergang von randomisierten Phase-IIb zu Phase-III Ansätzen kann mit entsprechenden statistischen Adjustierungen sehr gut implementiert werden.
Das Nutzen von historischen Kontrollen in die Studienplanung einer neuen Therapiestudie mit Weiterverwendung eines etablierten Standards kann nicht zur Verringerung der Fall verwendet werden.

Die Verwendung von historischen Kontrollen in der Fallzahlplanung und Analyse nachfolgender Therapiestudien ist insbesondere dann sinnvoll, wenn historische Kontrolle und Standardtherapiearm der laufenden Studie zu ähnlichen/gleichen Ergebnissen führt.

Die Verwendung historischer Kontrollen in laufenden Therapiestudien macht eine Randomisation laufender Therapiestudien überflüssig.

Historische Kontrollen können nie zur Fallzahlreduktion bei laufenden Studien herangezogen werden.