Design and Analysis of Umbrella and Basket Trials

Axel Benner and Christina Beisel

Division of Biostatistics, German Cancer Research Center (DKFZ)
Heidelberg, Germany

October 20, 2016
Personalized Medicine

(Providing) the right drug for the right patient at the right dose and time.

Sadée & Dai, 2005

A form of medicine that uses information about a person’s genes, proteins and environment to prevent, diagnose and treat disease.

National Cancer Institute
Traditionally, tumor histology determines (cytotoxic) treatment
Biomarkers gain more importance for selection of treatment strategies, e.g. by enrichment trials.

Challenge: With multiple targets based on multiple markers we are often close to the situation that we are faced with in rare diseases.
Biomarker-driven Clinical Trials

**Example:** Acute Myeloid Leukemia (AML)

Genotype-adapted clinical intervention trials of the German-Austrian AML study group (AMLSG)

![Genotype-driven Clinical Trials Diagram](attachment:diagram.png)

- **Genotype**:
  - APL [PML-RARA]
  - CBF-AML [KIT]
  - AML FLT3mut
  - AML NPM1mut
  - AML MLLrearr
  - Other subtypes, mainly high-risk

- **Treatment/Study**:
  - NAPOLEON
  - GIMEMA/AMLSG/SAL
  - +/- ATO-ATRA-Ida

- **AMLSG Studies**:
  - AMLSG 09-09
  - AMLSG 15-10
  - AMLSG 16-10
  - AMLSG 19-13
  - AMLSG 20-13
  - AMLSG 21-13
  - AMLSG 22-14
  - AMLSG 23-14
  - AMLSG 24-48 hrs

- **Other treatments**:
  - +/- Midostaurin
  - +/- Dasatinib
  - +/- Crenolanib
  - +/- ATRA
  - +/- ATO-ATRA-Ida
  - +/- Panobinostat

**AMLSG Personalized Medicine Approach**

**Screening**

- 24-48 hrs
Umbrella Trials

Enroll marker-defined cohorts in parallel under the "umbrella" of a specific histology or tumor type
The umbrella design focuses on a single tumor type or histology

The reason and rationale for the umbrella trial design first and foremost is to facilitate screening and accrual of patients

Primary features of the umbrella design

- It involves a group of two or more enrichment designs within the same protocol
- It allows for randomized comparisons
- It can have flexible biomarker cohorts
- It allows to add/drop biomarker subgroups
Umbrella Trials

Example: FOCUS4 Trial (Kaplan et al. JCO 2013) for Colorectal Cancer

Fig 1. Trial schema for FOCUS4. (*) The molecular cohorts are arranged in a hierarchy from left to right. For example, a patient with both a PIK3CA mutation and a KRAS mutation will be classified into the PIK3CA mutation cohort. CRC, colorectal cancer; EGFR, epidermal growth factor receptor; EREG, epiregulin; FFPE, formalin fixed, paraffin embedded; HER, human epidermal growth factor receptor; IHC, immunohistochemistry; MMR, mismatch repair; OS, overall survival; P, placebo; PFS, progression-free survival; Rx, treatment.
What about biomarker-negative patients?
Analysis

Potential benefit from inclusion of marker-negative patients:

- Collect data for retrospective biomarker identification
- Investigate prognostic properties of biomarkers
- Non-prognostic markers: Pool of standard-of-care arms
- Prognostic markers: Include biomarker-status as factor variable $B_i$ in Cox model
Comparison of models in different populations

- **Approach 1:** Separate models for biomarker 1 and biomarker 2
  \[
  \lambda_1 = \lambda_0_1 \exp(\beta_1 \times treat_1) \quad \text{(sample size: } n_1) \\
  \lambda_2 = \lambda_0_2 \exp(\beta_2 \times treat_2) \quad \text{(sample size: } n_2)
  \]

- **Approach 2:** Exclude biomarker negative patients
  \[
  \lambda = \lambda_0 \exp(\gamma_2 \times B_2 + \beta_1 \times treat_1 + \beta_2 \times treat_2) \quad \text{(sample size: } n_1 + n_2)
  \]

- **Approach 3:** Include biomarker-negative patients
  \[
  \lambda = \lambda_0 \exp(\gamma_1 \times B_1 + \gamma_2 \times B_2 + \beta_1 \times treat_1 + \beta_2 \times treat_2) \quad \text{(sample size: } n) \\
  \]
  where \( n = n_1 + n_2 + n_0 \)
Comparison: bias for $\beta_2$

Approach 1: $\lambda_2 = \lambda_0 \exp(\beta_2 \times \text{treat}_2)$

Approach 2: $\lambda = \lambda_0 \exp(\gamma_2 \times B_2 + \beta_1 \times \text{treat}_1 + \beta_2 \times \text{treat}_2)$

Approach 3: $\lambda = \lambda_0 \exp(\gamma_1 \times B_1 + \gamma_2 \times B_2 + \beta_1 \times \text{treat}_1 + \beta_2 \times \text{treat}_2)$

\[\begin{align*}
\lambda_0 &= 0.05, \quad \gamma_1 = \ln(0.5), \quad \gamma_2 = \ln(2) \\
\text{Population proportions: } (B_0, B_1, B_2): &2:1:1, \quad (10,000 \text{ Simulations})
\end{align*}\]
Comparison: standard error for $\beta_2$

**Approach 1:**
$$\lambda_2 = \lambda_0 \exp(\beta_2 \times \text{treat}_2)$$

**Approach 2:**
$$\lambda = \lambda_0 \exp(\gamma_2 \times B_2 + \beta_1 \times \text{treat}_1 + \beta_2 \times \text{treat}_2)$$

**Approach 3:**
$$\lambda = \lambda_0 \exp(\gamma_1 \times B_1 + \gamma_2 \times B_2 + \beta_1 \times \text{treat}_1 + \beta_2 \times \text{treat}_2)$$

$\lambda_0 = 0.05$, $\gamma_1 = \ln(0.5)$, $\gamma_2 = \ln(2)$

Population proportions: $(B_0, B_1, B_2): 2:1:1$, (10,000 Simulations)
Small sample size bias and Firth correction

- Maximum-likelihood methods not necessarily unbiased for finite samples

- Langner et al. (2003) investigated behavior of bias in relation to sample size for Cox regression
  - Bias depends on sample size, but also on baseline risk and treatment hazard rate

- Small sample size bias in simulation study

- Use Firth (1993) correction to reduce bias
Comparison: bias for $\beta_2$ with Firth correction

Approach 1: $\lambda_2 = \lambda_0 \exp(\beta_2 \times treat_2)$

Approach 2: $\lambda = \lambda_0 \exp(\gamma_2 \times B_2 + \beta_1 \times treat_1 + \beta_2 \times treat_2)$

Approach 3: $\lambda = \lambda_0 \exp(\gamma_1 \times B_1 + \gamma_2 \times B_2 + \beta_1 \times treat_1 + \beta_2 \times treat_2)$

$\lambda_0 = 0.05$, $\gamma_1 = \ln(0.5)$, $\gamma_2 = \ln(2)$

Population proportions: $(B_0, B_1, B_2): 2:1:1$, (10,000 Simulations)
Summary of results

For smaller sample sizes:

- Reduction of bias by using combined analysis (Approach 2)

- Even further reduction of bias by including of biomarker-negative patients (Approach 3)

- Additionally small improvements for standard errors

Approaches perform similar for larger sample sizes

Differences smaller when Firth correction is used
Histology-agnostic enrollment of marker-defined cohorts ("baskets")
Basket trials allow the study of multiple molecular subpopulations of different tumor or histologic types within one study.

Primary features of basket trials

- The design affords the flexibility to continually open and close arms of the study
- They can include highly rare cancers that would be difficult to study in randomized controlled trials
- Countless possibilities exist in designing and analysis of basket trials, such as writing protocols for each cohort and creating a screening and treatment infrastructure.
Basket Trials

Characteristics
- Marker-defined cohorts
- Typically non-randomized
- Primary purpose: treatment

Challenge Multiple targets → close to rare diseases trials
Example: CUSTOM Trial (Lopez-Chavez et al. JCO 2015)

From February 2011 to December 2012, 647 patients were enrolled onto the study. Of these, 569 patients (88%) had at least one molecular analysis attempted. Of these, 257 patients (45%) harbored a genetic abnormality in at least one of the core genes tested, and that was successfully performed. Of these, 212 patients (82%) were considered screen failures (Appendix Table A2, online only), and 45 patients were wild type or unknown for the mutations of interest (n = 313). Unsuccessful molecular profiling* (n = 77) could be evaluated for response and survival (n = 602). Enrolled in the NOS arm and received standard of care treatment or were enrolled in other clinical trials and followed prospectively until death.

Met CUSTOM general eligibility criteria and underwent molecular profiling (n = 647)

Had a successful molecular profiling* (n = 569)

Core mutations (n = 257; 23 with multiple mutations)

Had EGFR mutations (n = 90)

Had KRAS, HRAS, NRAS, or BRAF mutations (n = 110)

Had PTEN, Akt1, or PIK3CA mutations (n = 31)

Had ERBB2 mutations or amplifications (n = 15)

Had KIT or PDGFRA mutations or amplifications (n = 11)

Were wild type or unknown for the mutations of interest

Unsuccessful molecular profiling* (n = 77)

Screen Failures 212 were positive for at least one of the core mutations of interest but failed to enroll in treatment arms.

Received Erlotinib (n = 16)

Received Selumetinib (n = 11)

Received MK2206 (n = 7)

Received Lapatinib (n = 8)

Received Sunitinib (n = 3)

Could be evaluated for response and survival (n = 16)

Could be evaluated for response and survival (n = 10)

Could be evaluated for response and survival (n = 7)

Could be evaluated for response and survival (n = 7)

Could be evaluated for response and survival (n = 3)

Arm 1: NSCLC (n = 15)

Arm 2: SCLC (n = 0)

Arm 3: TM (n = 1)

Arm 4: NSCLC (n = 9)

Arm 5: SCLC (n = 1)

Arm 6: TM (n = 0)

Arm 7: NSCLC (n = 4)

Arm 8: SCLC (n = 2)

Arm 9: TM (n = 1)

Arm 10: NSCLC (n = 6)

Arm 11: SCLC (n = 1)

Arm 12: TM (n = 0)

Arm 13: NSCLC (n = 2)

Arm 14: SCLC (n = 0)

Arm 15: TM (n = 1)

Long-term follow-up

Fig 1. Flow diagram of patient population and treatment assignments. EGFR, epidermal growth factor receptor; NOS, not otherwise specified; NSCLC, non–small-cell lung cancer; PDGFRA, platelet-derived growth factor receptor alpha; SCLC, small-cell lung cancer; TM, thymic malignancy. (*) Successful molecular profiling was defined as having at least one core molecular analysis successfully performed.
Statistical Evaluation of the NCT MASTER basket trial

NCT MASTER
The MASTER (Molecularly Aided Stratification for Tumor Eradication Research) program:

Analysis of high-throughput diagnostics and histopathological evaluations to generate hypotheses for new targeted tumor therapies.
**NCT MASTER - Flow Chart**

**Jun 2016**  
- Patient Sample Asservation: $N = 535$  
- Sequencing: $N = 435$  
- Bioinformatic Analysis  
- Clinical Evaluation  
- Validation  
- Molecular Tumor Board*  

**Feb 2016**  
- Patient Sample Asservation: $N = 446$  
- Sequencing: $N = 359$  
- Bioinformatic Analysis  
- Clinical Evaluation  
- Validation  
- Molecular Tumor Board*  

**Recommendation: 65%**  
**Treatment: 25%**

**PI3K-AKT-mTOR**  
**RAF-MEK-ERK**  
**Tyrosine Kinases**  
**Developmental Pathways**  
**DDR Signaling**  
**Other**  
**Hypermutated/ImmunoTherapy**

NCT Heidelberg: Stefan Fröhling, Christoph Heining, Hanno Glimm, Stefan Gröschel, Peter Horak  
Claudia Scholl (Functional Genomics)  
DKTK (German Cancer Consortium) – München, Frankfurt/Mainz, Dresden, Essen/Düsseldorf, Freiburg, Berlin, Tübingen
Actual Study Design

Molecular Tumorboard

Prioritization of all options

Treatment recommendation and decision on basket / stratum

yes

no

Physician’s choice (no recommendation) Group 3

Treated cohort Group 1

Control cohort Group 2

Endpoints:
- Feasibility
- Efficacy: Objective response according to RECIST 1.1 criteria or disease stabilization for ≥ 6 months

Baskets / Strata for evaluation:

- PI3K - AKT - mTOR
- RAF - MEK - ERK
- Tyrosine Kinases
- Developmental Pathways
- DDR Signaling
- Other

Treatment administered?

yes

no
Causal Inference

Causal effect of treatment

- Randomized controlled trials are the gold standard for causal inference.

- Unfortunately they are not always feasible for a variety of reasons, including ethical concerns.

- Consequently, in such situations assessment of causal effects must be derived from non-randomized studies.
Causal Inference

NCT MASTER Basket Trial

- Individual recommendation of treatment
- May be associated with confounding

Possible methods against bias

1. Direct adjustment for confounding in regression models
   - Logistic regression

2. Propensity score methods
   - Propensity score: Conditional probability of treatment assignment given baseline characteristics (Rosenbaum & Rubin, 1983)
   - Optimal matching

3. Use of Instrumental variables
Causal Inference - DAG

Directed Acyclic Graph (DAG)

A graph where all edges are directed (doesnt contain bidirected dashed arcs denoting unobs. common causes/confounders) and which contains no cycles

We use DAGs to identify the causal structure of the data.

$Y$ outcome of interest (*Response*); $D \in \{0, 1\}$ binary *Treatment* indicator

$X$ observed characteristics; $U$ unobserved characteristics

**Interest:** Causal effect $D \rightarrow Y$

- What would happen to $Y$ if $D$ was changed externally (exogenously) from 0 to 1?
- **NOT:** Find the best fitting model for predicting $Y$
(Hypothesized) essential graph

Causal Inference - DAG (NCT MASTER)
Webpage:
www.nct-heidelberg.de/en/research/nct-master

Contact:
Study Office
Phone: +49 6221 5636253
E-mail: master@nct-heidelberg.de

Project Leader and coordinator of the program
Prof. Dr. med Hanno Glimm
Head Applied Stem Cell Biology, (Assistant) Medical Director Translational Oncology

Prof. Dr. Stefan Fröhling
Head Molecular and Cellular Oncology Translational Oncology

Dr. Daniela Richter
Scientific Coordinator NCT Precision Oncology Program
Finally: Points to Consider

Challenges:

- Strata of small size
- Strong heterogeneity

Study Design:

- Adaptation to refine, add and remove biomarker-treatment strategy combinations
  Allow to refine baskets, to add new baskets, to remove baskets.

Evaluation strategy:

- Success of trial vs. success of strata
  Use chain procedures starting with global null hypothesis of no effect
- Apply Firth correction