NCS 2018

Mouse clinical trials of N=1: Do we reduce too much?

Hannes-Friedrich Ulbrich

Research and
Early Development Statistics

October 5, 2018

Agenda

- // the project
- // PDX models
- // N=1 design, related designs and a paradigm shift
- // resampling and simulations
- // outlook for the project

disclaimer:

The data used for this presentation were generated within OncoTrack, an already finished IMI project where Bayer had been a consortium member. The experiments were solely conducted at one site of another OncoTrack consortium member.
Any (mis-) interpretation of the data are in full and only responsibility of myself.

introduction

// IMI project ITCC-P4
Innovative **T**herapies for **C**hildren with **C**ancer – **P**aediatric **P**reclinical **P**roof-of-concept **P**latform
(https://www.itccp4.eu/)

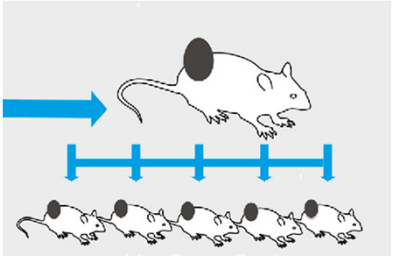
// public-private partnership supported by the 'Innovative Medicines Initiative' (IMI) of the EU.
// Bayer consortium member,
// among 12 academic institutions, 3 small-to-medium sized enterprises, 5 EFPIA members
// after an application period and signature of the Consortium Agreement
// project start: 2017
// duration: 5 years

// Description of Action (DoA) document – subsequent to Consortium Agreement:
// "n=1 study design" – a set feature for the to be planned PDX model experiments

PDX – patient derived xenograft

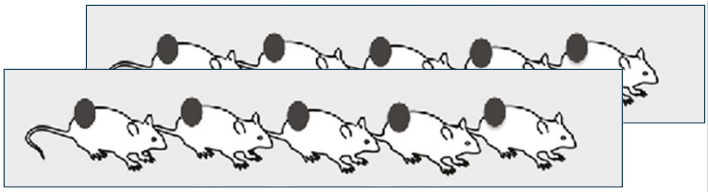
// starts with a part of a solid tumor (biopsy, …) of a particular patient
// sliced into pieces [there are other methods of split-up as well]
// pieces implanted into (special: immunocompromised, …) mice ➔ growing there – F0 generation
// if tumor burden too high: resection (removal of necrotic tissues, …)
// sliced into pieces ➔ implanted ➔ growing – F1
// and so forth up to Fn generation – the "magic" multiplication of tumor material
// many "grand-grand-…-children" of the patient's tumor
// biological question: with increasing generation
genetic and/or histological divergence from patient's tumor
➔ a question about the shift
// statisticians' question: increased variability

// these derivatives of the original tumor form **1 PDX model**



PDX experiments

- starts with – one or more – PDX models with its Fn generation pieces
  - implanted into mice
  - for getting treated



- implantations might be
  - heterotopic: implantation independent of tumor origin's location – usually subcutaneous
  - orthotopic: implantation into the corresponding anatomical position, e.g. brain tumor PDX model into the mouse brain
- measuring tumor growth – might become more complicated for orthotopic than heterotopic PDX model applications

PDX and the N=1 design

- as given in ITCC-P4 consortium documents
  - different tumor types and subtypes – particularly important to children
  - ~20 PDX models to be used for each type – most heterotopic = subcutaneous, brain: orthotopic
  - N=1 design –
    - to be used for each PDX model
    - specified as: 3 mice getting (a unique) control + 1 mouse per treatment
    - joint consideration of the 20 PDX models
    - no statistical model given
- seems mainly to be based on ideas specified in a competitor's internal poster presentation
  - about 70 PDX models used (per indication / substance set testing)
  - tumor growth and time-to-event approach

PDX and the N=1 design

- design stage – work in progress, current status
  - experiment level: aiming for 20 PDX model replicates
  - patient level (PDX model): 11 (3 + 8*1) replicates [current pressure: increase to 9*1]
  - treatment level: 3 (reference group), 1 per treatment/substance

  here: reference group = either placebo or a vehicle [vehicle to 1 particular substance]

- statistically speaking:
  - 20 PDX models are basis for an experiment
  - hierarchically subdivided into the levels as above
- experimental conduct perspective:
  - 20 separate experimental conducts (20 PDX models realized)
  - each realized PDX model uses **1 mouse per** treatment/substance – the **N=1** design

'classical' PDX experimentation

- more common until recently:
  one PDX model – one (large) experiment
  - 6 mice (e.g.) per control and treatment group
  - looking for substance/treatment differences … within one PDX model!
  - aiming for associations with genetic/genomic properties of the tumor
  - (one of) the challenge(s): to have enough mice with growing tumor @ treatment start (randomization)
    - often not feasible ➔ change in design
    - most common: "pragmatic changes"
      - "staggered" start
      - experimental split – some substances now, some later
      - BUT: randomized?
- as done in IMI OncoTrack ➔ data base for my undertaking

Experimental use of PDX models – recent paradigm shift

- instead of using 1 PDX model ➔ experiment on many PDX models (of the same type)
  - reflecting the increasing number of PDX models available (e.g., @ specialized CRO)
  - better mimicking the patient focus of later development – translational aspect

    "Increase the power of translational research using predictive PDX cohorts reflecting the human tumor heterogeneity and diversity (each model represents 1 patient) …" (AACR meeting poster)

- interestingly:
  doing the same with replication on the mouse level, i.e.,
  having at least two mice per PDX model and substance/treatment is
  - mentioned as a possibility by statisticians
  - off focus (?) by biologists
  - explained: although within limits of feasibility (i.e., a constant upper number of mice to be handled)
    - less PDX models would not reduce costs
    - less substances is … out of imagination(?)

Experimental use of PDX models – recent paradigm shift

- as often with new (hot) topics and paradigm shifts – terminology is an issue
  - to separate from "the old"
  - to describe the new as good as possible in an rapidly evolving field

- here:
  - mouse clinical trial
  - 1 × 1 × 1 approach/design
    (occasionally translated into 1 PDX model / 1 mouse / 1 treatment – a rather confusing explanation)
  - PDX surrogate clinical trial
  - single mouse preclinical trial
  - N=1 design

  are – my guess – meaning the same or – at least – something very similar

  some terminology triggered by (the current) PDX model = patient = statistical unit

How to evaluate any positive effect on PDX models?

- time-to-event approach
  - mentioned on posters as having been applied successfully, but explained in rather vague manner (e.g., TTP = time to tumor progression without explaining how to get these consistently)
  - with 20 PDX planned – not the way to go (?)
- tumor growth evaluation
  - somewhat classical: T/C (treatment over controls) ratio
    - (commonly) after-the-fact determination of suitable end-timepoint (no drop-outs yet)
    - 'cross-sectional' judgement – beyond-chosen-timepoint measurements not taken into account
    - based on arithmetic means – additive model
  - relative change from baseline (aka percent change in tumor volume)

  objections to these (and others) from both biologists and statisticians

tumor growth evaluation

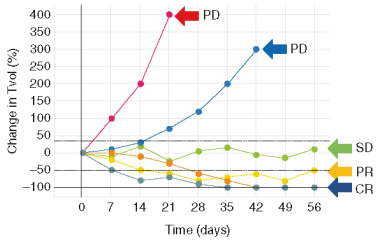
- objections ➔ newer approaches
  - biologists: RECIST
  - statisticians: longitudinal modeling, taking care of missing information
- RECIST (response evaluation criteria in solid tumors)
  - established 2000 (updated 2009, 2016) for individual patients
  - four levels: complete response (CR), partial r. (PR), stable disease (SD), progressive d. (PD)
  - adaptations for PDX model becoming fashionable
    - 4 categories – using same terminology
    - different definitions out there

tumor growth evaluation – (m)RECIST

mouse-individual judgements [Williams JA (2017)] ← → treatment group based judgements [Gao H, et al (2015)]

Table 10.1 Mouse clinical trial assessment of tumor response to treatment with sensitivity parameters based on RECIST

Clinical term	Clinical	Preclinical
Complete response (CR)	Disappearance	Complete regression
Partial response (PR)	>30% decrease in the sum of longest diameters	>50% decrease in tumor volume
Stable disease (SD)	<30% decrease and <20% increase in the sum of longest diameters	<50% decrease and <35% increase in tumor volume
Progressive disease (PD)	>20% increase in sum of longest diameters	>35% increase in tumor volume



Response is assessed by tumor volume change compared to the baseline tumor volume:
$\%\ TVol = 100 \times (V_{max} - V_{initial}) / V_{initial}$

- Determination Steps
  - V = percent change in tumor volume for each animal (end_volume - start_volume) / start_volume) * 100
  - V_m = within group minimum of V
  - V_a = (arithmetic) group mean of V
  - Determine RECIST category according to the table

RECIST Category	Best Response: V_m	Average Response: V_a
Complete Response (CR)	< -95%	< -40%
Partial Response (PR)	< -50%	<-20%
Stable Disease (SD)	< 35%	< 30%
Progressive Disease (PD)	Anything else	

statistical model

- tumor growth per mouse [Choudhury KS, et al (2012), Hather G, et al. (2014)]
  - exponential – assuming
    - (almost) no necrosis
    - constant tumor growth rate from treatment start on – immediate treatment effect
  - early drop-out (exclusion) because of to heavy tumor burden – modeled by beyond observation prediction
- "derived measure": tumor growth rate per subject
  ["giving a value even if the assumptions were not fully met", Hather et al]
- can be "averaged" per group
- further (common) measures derivable:
  - "model based" T/C – @ across experiments comparable timepoints
  - "model based" tumor doubling time (tumor halving time)
  - "model based" RECIST – @ across experiments comparable timepoints

simulations: selection and model

- OncoTrack data base
  - colon cancer – primary tumors only (i.e., no metastases) of adults
  - subcutaneous [heterotopic] implantation
  - conducted as 1 experiment per PDX (often split into sub-experiments with sub-ex specific ref group)
  - @ one site, experiments (and sub-ex) spread over calendar time
  - 7 most consistently available substances/treatments
- here: taken as basis for resampling based modeling

simulations: selection and model

- OncoTrack data base
  - 41 PDX models (of about 200) – valid according criteria [primary tumor, colon tumor]
  - 13900 records, i.e., tumor size response values
  - 1740 mice involved,
    - ~ 12.5 values per mouse = timepoints
    - all with baseline value (time=0)
  - different timepoint patterns
    - observation phases PDX model dependent: between 15 and 32 weeks
    - post treatment start observational timepoints: min week 2, max week 32 (never @ 1, 30, 31)
    - most common timepoint: week 3 (75% of all mice)
    - least common: weeks 19 & 26 (less than 2% of all mice)

simulations: selection and model

- OncoTrack ➔ two level sampling:
  - PDX model sampling: 20 out of the 41 [also 10 out of 41] – repeated 26 times
  - animal level sampling: 3 of placebos, 1 of each of 7 substances – repeated 100 times
- evaluated the 'full' model (6 animals per group)
  the 100 animal sampling level models
  - $\log_{10}$(tumor volume) = time time*substance + Gaussian error
    - substance = fixed effect, time = covariate
    - relying on randomization assumption on time=0
    - error increasing with tumor size
  - random effects: PDX model (Gaussian on log scale)
  - mouse cluster: autoregressive repeated structure
  - SAS® proc GLIMMIX

simulations: selection and model

- SAS® proc GLIMMIX
  - non-convergence?
  - "loop" through different optimization techniques
    [2 different Newton-Raphson, quasi-Newton, conjugate-gradient, double-dogleg, Nelder-Mead simplex, trust-region]
    - until convergence is reached
    - keeping unsuccessful optimizations on record
- the main questions:
  - first level:
    how much do we loose by having only 1 mouse per treatment (3 controls) per PDX model compared to having 6
  - second level:
    if there were remarkable differences, how much would it help to take 3 mice per treatment-PDX

simulations: results

'*But lemma 1 was wrong.*'

Helga Königsdorf: Lemma I. In: Meine ungehörigen Träume, 1981
republished 2016 in Mitteilungen der DMV [quarterly bulletin of the German Mathematical Society]

- translate 'lemma 1' into 'PDX model sampling' … ➔ you get the dilemma
- i.e., simulation results will be presented in a future presentation

how to judge?

- statistically: sampling distribution "around" the 'full' model (6 mice per group)
- biologically?:
  - categorizing:
    - increasing / shrinking
    - growth rates: shrinking / steady state / moderately growing / remarkably growing – thresholds?
  - RECIST classification @ time point

Summary

- a strategy to utilize a former IMI project's data
  to give design recommendations for upcoming PDX model experiments
- whether adult PDX models in colon cancer can be relevant for discussing children's tumors
  must remain open
- according to the literature
  - the '1 × 1 × 1 approach' in 'mouse clinical trials' is considered to be successful,
  - published 'proofs' of that claim are outstanding (i.e., I haven't seen any, yet)
- any statisticians' ideas (recommendations?) to try with at least 2 mice per …
  have still to be published (and conducted before)
- the current IMI project's (ITCC-P4) approach is slightly different:
  it's intended to use 3 mice for the control group

outlook: design bound by reality

- the current IMI project's (ITCC-P4) approach goes for
  - 1 mouse per PDX model and treatment/substance
  - 3 mice for the control group
- as it turns out: this might have been – even for the statistician – a great idea
  - the specialists in the field of child cancer PDX models are in serious doubt that implanting a PDX model into 3+8 (or 9) mice at one particular day would lead to similar initial tumor growth
    - contradicts the idea of randomize/treatment start all 3+8 on the same day
    - it's expected that they will have to split into 2 or 3 different start date batches per PDX model
  - having the chance to randomize 1 control to each start date batch might become partial relief

references

- Choudhury KS, et al (2012). A comparison of least squares and conditional maximum likelihood estimators under volume endpoint censoring in tumor growth experiments. Stat Med 31: 4061–4073.
- Gao H, et al (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nature Medicine 21: 1318-1328.
- Hather G, et al (2014). Growth rate analysis and efficient experimental design for tumor xenograft studies. Cancer Inform 13 (Suppl 4): 65-72.
- Williams JA (2017). Patient-Derived Xenografts as Cancer Models for Preclinical Drug Screening. doi: 10.1007/978-3-319-55825-7_10
- Zhu AZX (2018). Quantitative translational modeling to facilitate preclinical to clinical efficacy & toxicity translation in oncology. Future Sci OA (2018) 4(5), FSO306. doi: 10.4155/fsoa-2017-0152

further sources:

- doi :10.1016/j.celrep.2017.10.021
- poster 2170 of AACR Annual Meeting 2018

Thank you!

hannes-friedrich.ulbrich@bayer.com