**2.6 Clinical Trials: Current + NextGen**

Alihan Huyuk

[ah2075@cam.ac.uk](mailto:ah2075@cam.ac.uk)

**Clinical trials**

* Difficult, expensive, and ripe for disruption
* Average PoS per Phase II/III trial: 50-60%
* 1-2 years for trial
* Up to 2,000 trials per year
  + Half for oncology
* Opportunity
  + Despite flaws, RCTs remained dominant approach for decades
  + Innovation:
    - Improve patient benefit
    - Reduce required patient numbers
      * Replace controls and preventing dropouts
    - Reduce cost per patient
      * Improved recruitment, monitoring and operations
    - Improve PoS
      * Better recruitment, prediction, trial design, and analytics
  + Cost and time savings by running better trials
* Clinical trial phases

|  | Exploratory | | Confirmatory | Extrapolatory |
| --- | --- | --- | --- | --- |
|  | Phase I | Phase 2 | Phase III | Phase IV |
| Goal | Safety | Dose finding | Efficacy | Retargeting |
| # of patients | 10 -100 | ~100 | ~1000 | >1000 |
| Regulations | Less regulated, company’s own risk | | Highly regulated | Regulated |
| Input data | * Pre-clinical (Animal models) * Related clinical trials * Baseline data (current practice) * PKPD models | | * Results from Phase I/II * Interim analysis | * Real world usage (incl. Off-label data & other drugs) |

* Steps taken for each phase of the study
  + Planning
    - Spec out the trial protocol
  + Conduct
    - Logistical challenges
  + Analysis
    - Produce trial findings
  + Clinical Use (for Phase IV)
    - Usage of the approved treatment
* Challenges and potential solutions felt through the clinical development journey
  + Step 1: Planning -> what data?
    - Causal ML
      * Determining endpoints
    - Transfer Learning
      * Incorporating observational data
      * Incorporating pre-clinical data
      * Incorporating cross-trial results
    - Optimal Design
      * Optimal design of trials
    - Synthetic Data
      * Augmenting preclinical/ cross-trial data
  + Step 2: Conduct
    - Reinforcement learning/ Multi-armed Bandits
      * Determining dosage
      * Discovering drug combinations
      * Identifying “good” subpopulations
      * Recruitment “right” patients
    - Inverse Reinforcement Learning
      * Retaining recruited patients
    - Synthetic data
      * Streamlining data sharing
        + internal/ external
        + Contractors
  + Step 3: Analysis
    - Treatment effect estimation
      * Predicting personalized response
      * Subpopulation analysis
    - Time series analysis
      * Time to event analysis
      * Competing risk analysis
    - Few-Shot learning
      * Combining models (e.g. PKPD)
    - Synthetic data
      * Anonymising results for reporting
  + Step 4: Clinical use
    - Treatment effect estimation
      * Refining guidelines (e.g. timing, dosage)
      * Indication expansion
    - Time series analysis
      * Effect and impact on comorbidity
      * Modeling disease progression
    - Few-shot learning
      * Rapid deployment
    - Synthetic Data
      * Facilitating RWD access and anlayiss
* Machine learning can help in two ways:
  + SyncTwin
    - <https://proceedings.neurips.cc/paper/2021/hash/19485224d128528da1602ca47383f078-Abstract.html>

**Adaptive Recruitment**

* Identifying good subpopulations fast with confidence

AdaGGI:

<https://arxiv.org/pdf/2208.05844.pdf>

Individualized treatment effect inference:

<https://www.vanderschaar-lab.com/individualized-treatment-effect-inference/>

Next generation clinical trial:

<https://www.vanderschaar-lab.com/adaptive-clinical-trials/>