Topics in Machine Learning
Machine Learning for Healthcare

Rahul G. Krishnan
Assistant Professor
Computer science & Laboratory Medicine and Pathobiology

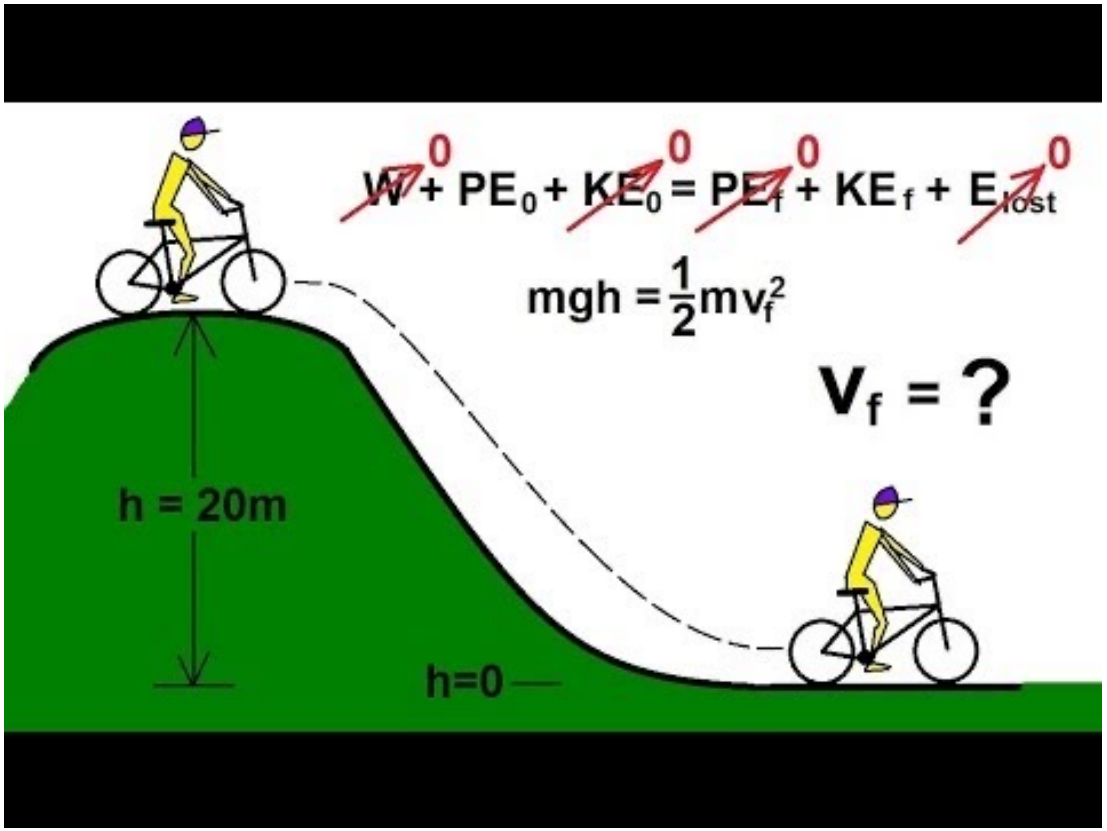
Announcements

- Most (if not all) of you have submitted your paper summary assignments
- Next up – in ~3-4 weeks you will begin presenting on your projects
  - Get started early
  - Book TA time
  - Come to office hours
- Mid-quarter feedback form – Please do fill out ASAP.

Outline

- Decision making with mechanistic knowledge
- History of decision making for interventions in healthcare
- Randomized control trials
  - Cohort design
  - Phases
  - Controls
  - Blinding
- Challenges in RCTs and approaches to mitigate it

Interventions in physical systems



Equations characterize our understanding of physical systems

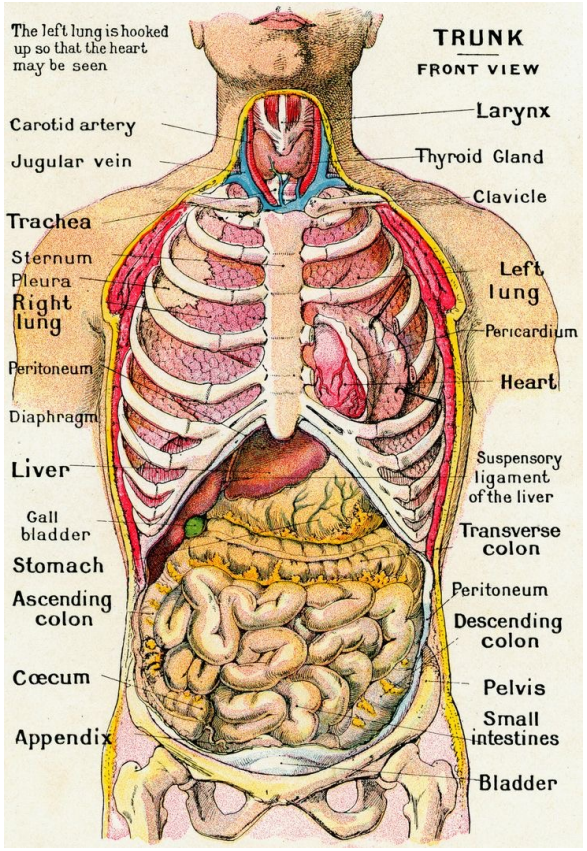
Laws of conservation give us equalities that must hold

We can use these equations to answer questions around the effect of interventions

Source: https://www.youtube.com/watch?v=yMdSrTyXuKU

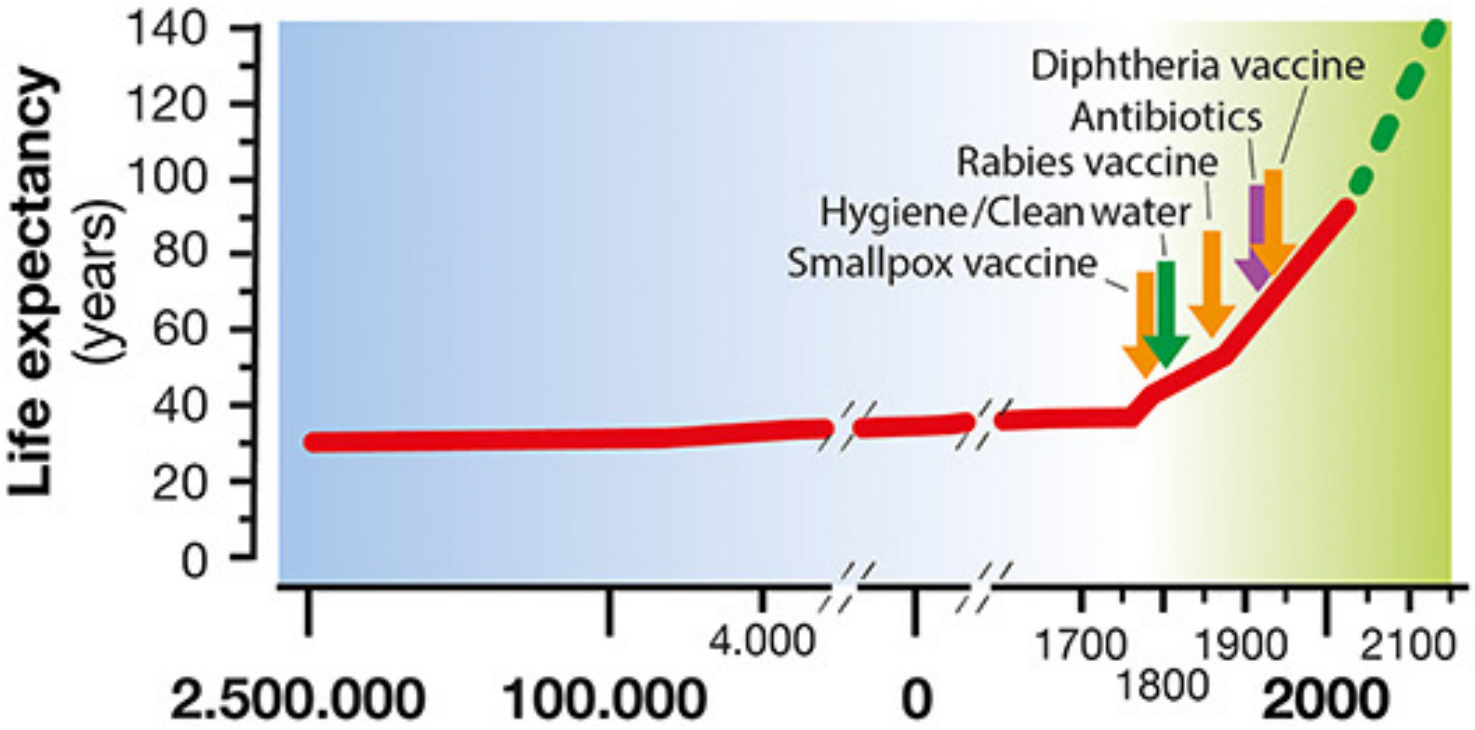
Medicine has progressed a long way

- We've come a long way in our understanding of the human body
- But processes in the body occur at multiple scales and we lack mechanistic models to characterize the effect of every intervention



How should we make decisions in healthcare?

History



Rosini et. al, Vaccines Against Antimicrobial Resistance, 2020

Discovery of early medicines/interventions

- Penicillin
  - Discovered in 1928 by accident when studying the properties of the *Penicillium* mold
  - Hypothesized mechanism was unknown but scientists knew it killed bacteria
  - Reason for use:
    - First used to successfully treat an eye infection in children
- 1945 – First Randomized Control Trial
  - Published "Streptomycin treatment of pulmonary tuberculosis" by Austin Bradford Hill
  - May be likely that the idea of randomization existed long before medical publications
  - Prospective biomedical research study designed to answer specific questions about an intervention

What is a randomized control trial?

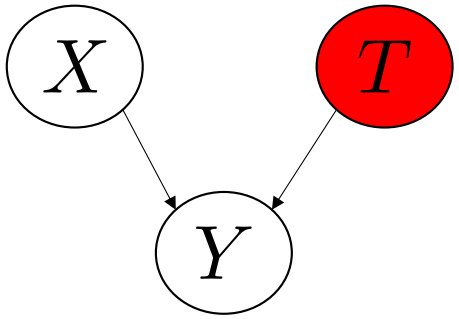
Control: Blue pill [the treatment representing the standard of care]

Treatment: The new intervention that is being evaluated

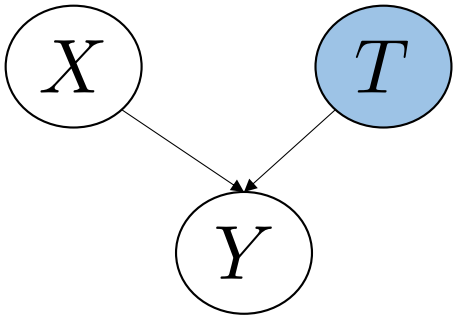
Randomize control trials

- Step 1: Select a cohort of individuals
- Step 2: Randomly split the cohort into two groups
- Step 3: Give one group the control and give the other group the treatment
- Step 4: Observe the patients over time and see who got better

Randomization as a graphical model



Treatment



Control

$$ATE = \sum_{i \in \mathcal{T}} y_t - \sum_{i \in \mathcal{C}} y_c \approxeq \mathbb{E}[Y_t - Y_c]$$

Y is an outcome of interest (positive value good; negative value bad)

Features of a randomized control trial

- Cohort design & randomization
- Study phases
- Controls
- Blinding

Cohort design

Cohort Design & Study plan

- Selection criteria: Who qualifies to participate
- How many people in study?
- Length of study & choice of intervention
- **Discussion: If you were a drug company, and were designing the cohort, what are considerations you might want to incorporate in the design of your cohort.**

Why does randomization work?

- Recall Mike Fralick's talk
- There are many demographic factors that are shared/unique across patients
- Randomization ensures that these are evenly distributed across treatment and control groups
- Ensures that there is no *selection bias* in the cohort

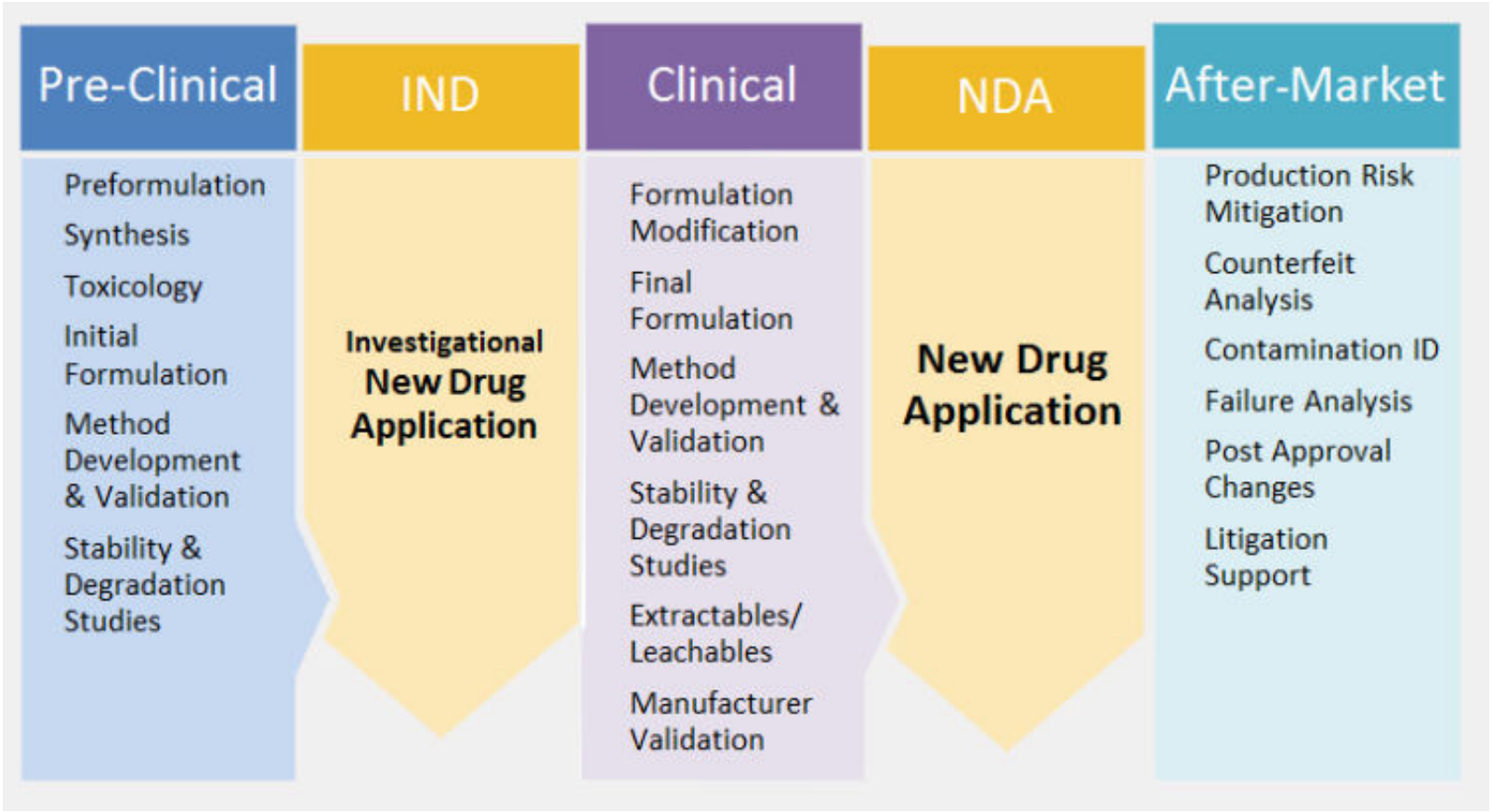
Types of randomization

- Complete randomization
  - Flip a coin
- Stratified randomization
  - Ensuring balance across imbalanced covariates in treatment/control group
  - 100 people but only 10 males in cohort
- Cluster randomization
  - What if you want to test an intervention across “groups” rather than individuals
  - Useful if you suspect the treatment effect across individuals in a group is correlated
  - Examples:
    - Clustering children into classrooms and applying randomization at the classroom level

Phases of a clinical trial

Phases of modern drug design

- Develop a new drug
- Preclinical phase:
  - Initial testing
- Phase 1: Small scale
  - <30 people
  - Goal: Finding a minimum safe dosage for the drug
- Phase 2: Medium scale
  - <100 people
  - Goal: Assessing initial benefit for the drug
- Phase 3: Large scale
  - 100-1000 people
  - Goal: Assessing benefit relative to current standard of care


Pre-Clinical
IND
Clinical
NDA
After-Market
Preformulation
Synthesis
Toxicology
Initial Formulation
Method Development & Validation
Stability & Degradation Studies
Investigational New Drug Application
Formulation Modification
Final Formulation
Method Development & Validation
Stability & Degradation Studies
Extractables/ Leachables
Manufacturer Validation
New Drug Application
Production Risk Mitigation
Counterfeit Analysis
Contamination ID
Failure Analysis
Post Approval Changes
Litigation Support

Controls

Controls

- Controls are the *alternative* that is being considered to the treatment that is proposed
- No treatment
  - Common in new surgical procedures
- Placebo
  - Duplicate the experience of the intervention without its effect
- Standard of care
  - Comparison to current clinical practice – usually in Phase 3

Blinding

What is blinding?

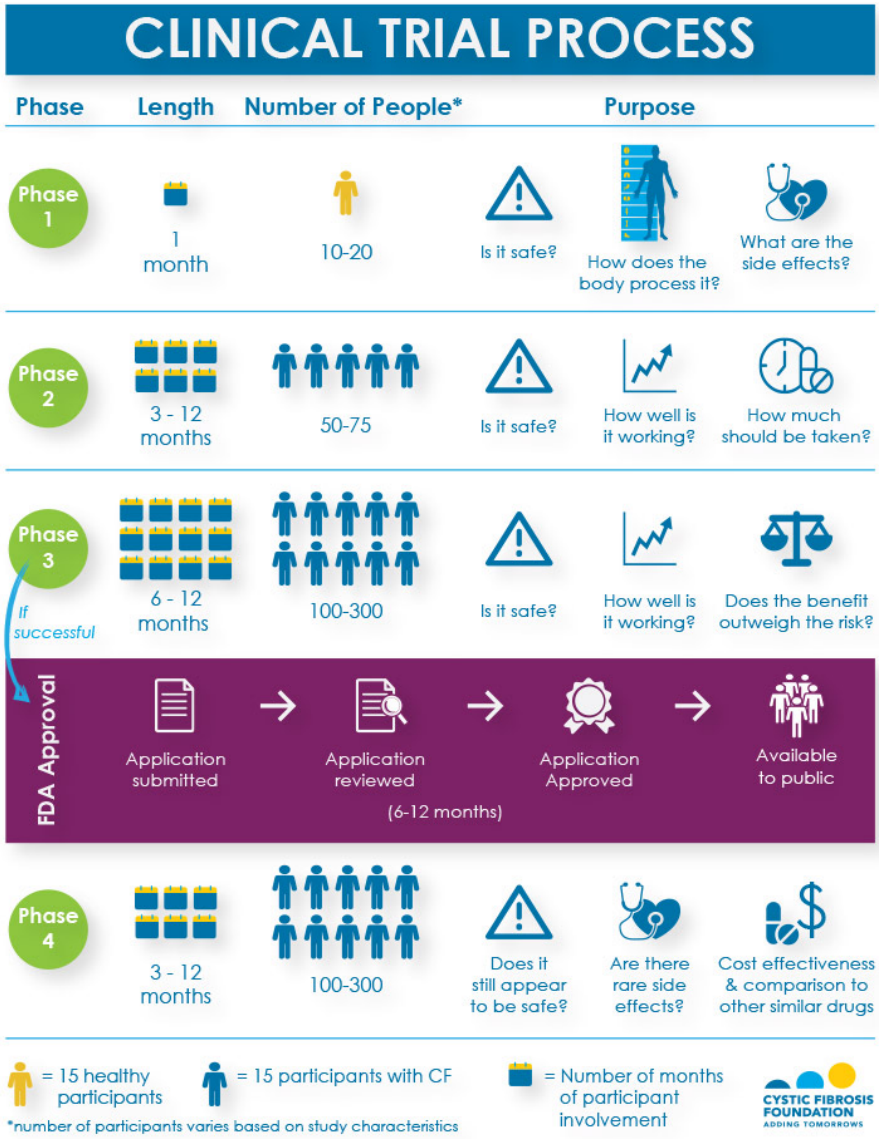
- Blinding:
  - Hiding the identity of who received treatment and who received the control
  - Prevent unintentional bias
- **Discussion: Why do we need to have blinded trials?**

Types of blinding

- Open label
  - Participants and doctors know if they are on treatment/control
- Single
  - Doctors know but participants do not know if they are on treatment/control
- Double
  - Conceal nature of treatment from participants and doctors
  - Only those directing the study know
- Triple
  - Conceal nature of treatment from participants, researchers and administrators/doctors

Challenges in designing and running RCTs

- Cohort selection
  - Need consent and ways to find people who meet the criteria for being part of an RCT
  - Why is it hard: Healthcare infrastructure is disperse
- Cost
  - Multiple stages of the RCT are costly
  - Why is it hard: Require large groups of patients, administrators and coordination between organizations (drug companies and hospitals)
- Time
  - The end-to-end pipeline for drug development takes a *long* time



Discussion: Can we use machine learning to help? If so, where and how?

Machine learning to evaluate eligibility criteria

- Evaluating eligibility criteria of oncology trials using real-world data and AI, Liu et. al, Nature Medicine 2020
- Problem being tackled:
  - Recruitment for trials is challenging
  - Trials can have low enrolment (86% of trials fail to meet recruitment within time)
  - **Why**: Eligibility criteria can be too strict but not all clinicians agree on how to relax them
- **Key idea:** Can we use real-world electronic health record data to emulate clinical trials and use the data to *relax* eligibility criteria.

Dataset

- [a] Flatiron Health EHR-derived database (private database)
  - De-identified data from 280 cancer clinics
  - Cohort: [aSCLC] Advanced small cell lung carcinomas [~61K patients)
- [b] ClinicalTrials.gov – contains information on trials and their associated eligibility criteria
- Took criteria from [b], encoded them as rules in [a] and selected patients who *would have met the criteria for 10 aSCLC trials*

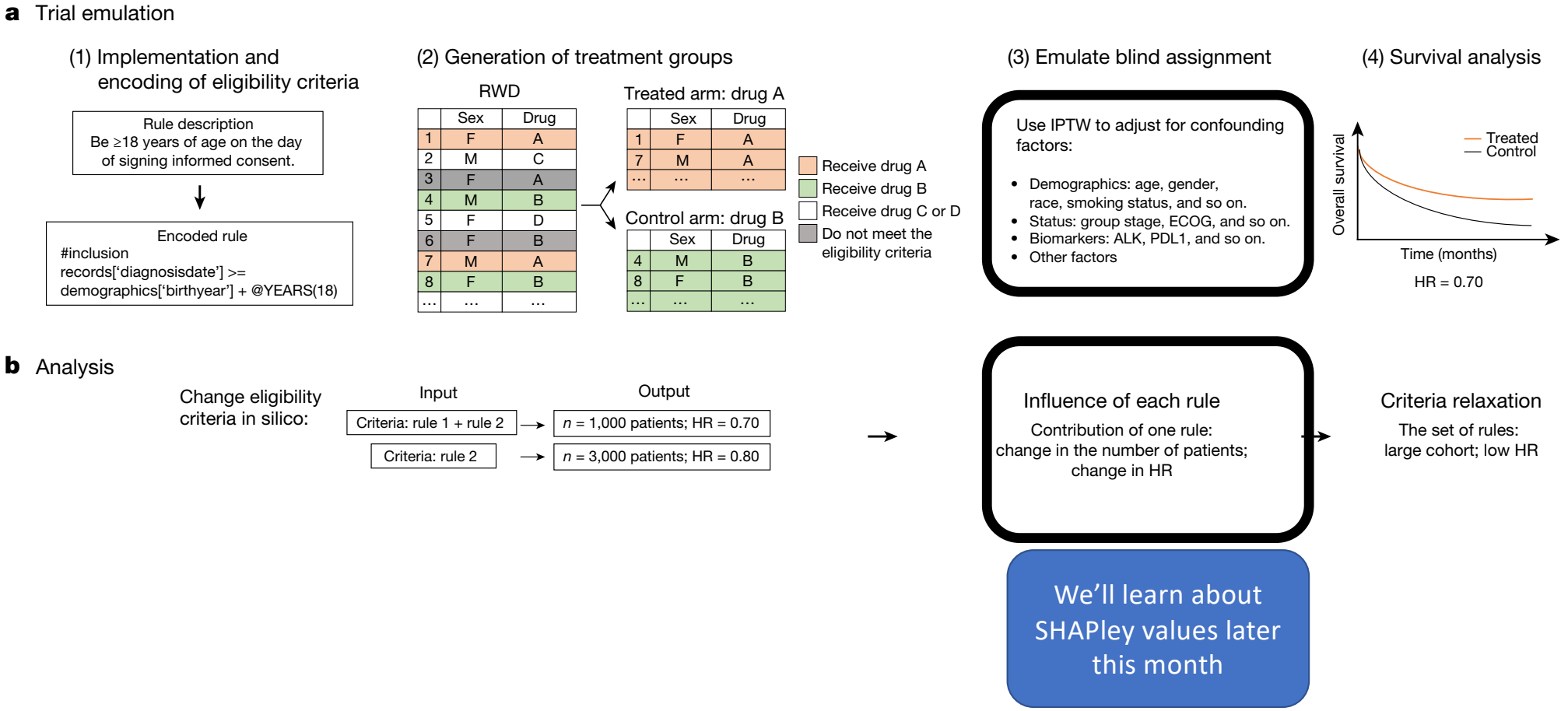
Table 1 | Comparisons of eligibility criteria

Trial name	Original trial criteria		
	No. of criteria	No. of patients	HR
FLAURA	10	2,277	0.81
LUX8	11	129	0.65
Checkmate017	17	523	0.67
Checkmate057	19	792	0.75
Checkmate078	18	1,509	0.74
Keynote010	13	806	0.56
Keynote189	15	4,066	0.88
Keynote407	13	2,031	1.13
BEYOND	12	2,902	1.09
OAK	19	493	0.88
Average	15	1,553	0.82

Study finds that hazard ratios for certain treatments comparable to study criteria in the entire patient cohort

Suggests that there is potential benefit to developing relaxed study design protocols

Pipeline in a slide



Key insights - (1)

Table 1 | Comparisons of eligibility criteria

Trial name	Original trial criteria			Fully relaxed criteria		Data-driven criteria		
	No. of criteria	No. of patients	HR	No. of patients	HR	No. of criteria	No. of patients	HR
FLAURA	10	2,277	0.81	3,819	0.82	4	2,546	0.75
LUX8	11	129	0.65	1,350	0.81	5	141	0.58
Checkmate017	17	523	0.67	4,900	0.71	7	4,085	0.71
Checkmate057	19	792	0.75	4,900	0.71	9	2,594	0.66
Checkmate078	18	1,509	0.74	4,900	0.71	9	3,348	0.68
Keynote010	13	806	0.56	1,950	0.51	1	1,948	0.51
Keynote189	15	4,066	0.88	8,818	0.94	7	4,595	0.85
Keynote407	13	2,031	1.13	10,437	1.07	4	9,173	1.04
BEYOND	12	2,902	1.09	9,310	1.14	4	3,043	1.08
OAK	19	493	0.88	1,288	0.87	6	620	0.80
Average	15	1,553	0.82	5,167	0.83	6	3,209	0.77

The number of inclusion/exclusion criteria, the number of eligible patients and the hazard ratio of the overall survival of emulated aNSCLC trials with eligibility criteria under three scenarios: the original criteria used in the trial, fully relaxed criteria and data-driven criteria. The fully relaxed criteria correspond to evaluating the hazard ratio of the overall survival of all of the patients in the Flatiron database who took the treatments in the relevant line of therapy. The data-driven criteria were selected by Shapley values. HR, hazard ratio.

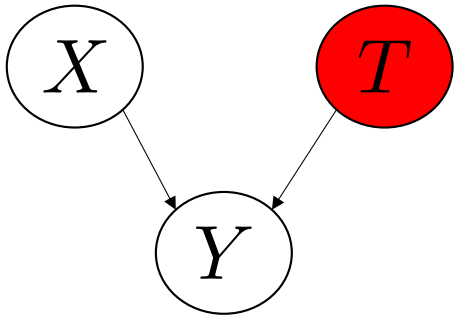
Key insights – (2)

- More inclusive trial resulted in more women being selected into the trial cohort
- Expanded data-driven eligibility included older individuals (without limiting comorbidities being selected into the trial)

Summary

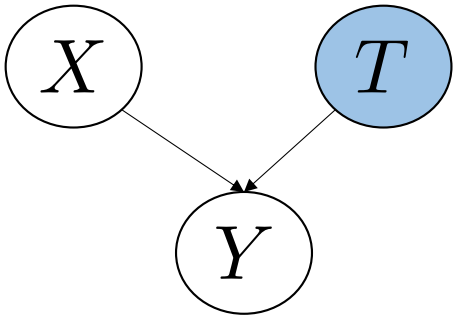
- Hearsay: “Only ~10-15% of interventions in surgical specialties are backed by evidence from randomized control trials”
- Using observational data holds a lot of promise for the design of effective mechanisms
- But there is enormous debate in the community:
  - The Magic of Randomization versus the Myth of Real-World Evidence, Collins et. al, 2020

Next week



Treatment

How (and when) can we use observational data to mimic an RCT.



Control

$$ATE = \sum_{i \in \mathcal{T}} y_t - \sum_{i \in \mathcal{C}} y_c \approxeq \mathbb{E}[Y_t - Y_c]$$

Y is an outcome of interest (positive value good; negative value bad)