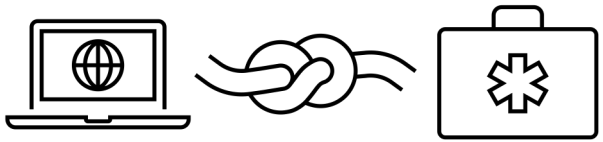
# The role of machine learning in clinical research: transforming the future of evidence generation

Yuxiao (Shawn) Sun  Yuyi (Taylor) Ding

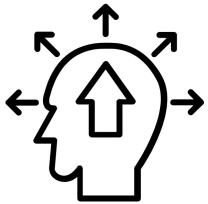


# 1.
# Background

*Improved Computing Resources*

*Data Availability*

*Novel Methods*



# CONTENTS

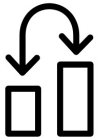
# 2.

# Preclinical drug discovery and development research

# 2.1 Target drug discovery

Analysis existing research

Generate molecules



Different drug performance

# 2.1 Target drug discovery

- **Obsessive-compulsive disorder (OCD) drug[1]**
  - 1/3 cost
  - 12 months vs. 5 *years*
  - *350 compounds vs. 2500 compounds*

*[1]: Koh D. Sumitomo Dainippon Pharma and Exscientia achieve breakthrough in AI drug discovery: Healthcare IT News - Portland, ME: Healthcare IT News; 2020.*

# 2.2 Clinical trials planning

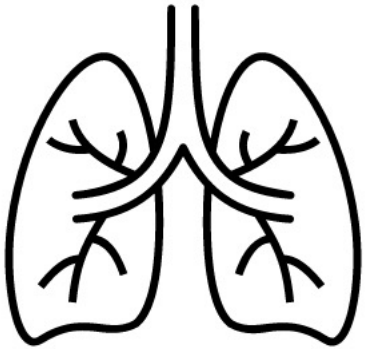
- Maximizing the success and efficiency of trials

Planning phase

Optimize the choice of treatment regimens

# 2.2 Clinical trials planning

- Reinforcement Learning for clinical trials in nonsmall cell lung cancer(NSCLC) [1]



Discover optimal treatment

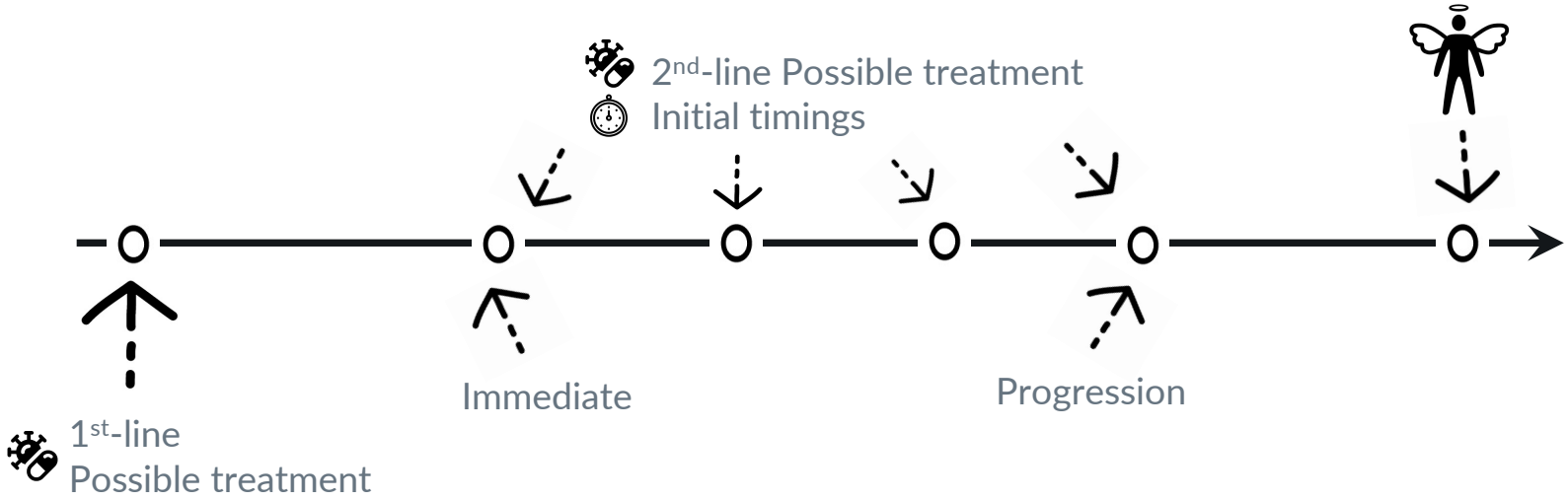
Q – learning framework [2]

*[1]: Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. Biometrics. 2011; 67(4):1422–33. https://doi.org/10.1111/j.1541-0420.2011.01572.x.*

*[2]: Watkins, C. J. C. H. (1989). Learning From Delayed Rewards. Ph.D. Thesis, King's College, Cambridge, U.K.*

# 2.2 Clinical trials planning

- Treatment plan and therapy options[1]




*[1]:* Socinski, M. A. and Stinchcombe, T. E. (2007). Duration of first-line chemotherapy in advanced nonsmall-cell lung cancer: Less is more in the era of effective subsequent therapies. Journal of Clinical Oncology 25, 5155–5157.

# 2.2 Clinical trials planning

- Primary goal

Optimal Compounds

Optimal time to initiate 2nd – line therapy

# 2.2 Clinical trials planning

- In the clinical setting – Q-learning

State $S_t$   Patient covariates and treatment history

Action $A_t$    Treatment options and timing

Reward $R_t$   Survival time

## 2.2 Clinical trials planning

- Q - learning

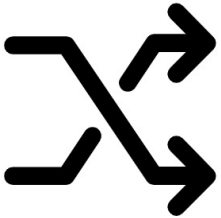
$$Q_t(s_t, a_t) = E\left[ R_t + \max_{a_{t+1}} Q_{t+1}(S_{t+1}, a_{t+1}) \middle| S_t = s_t, A_t = a_t \right].$$

# 2.2 Clinical trials planning

- Limitations

⚠ Assumption :
Survive → 2nd – line therapy

⚠ Peer review



**Conceptual Promise**

# 3.

# The role of ML in clinical trial participant management

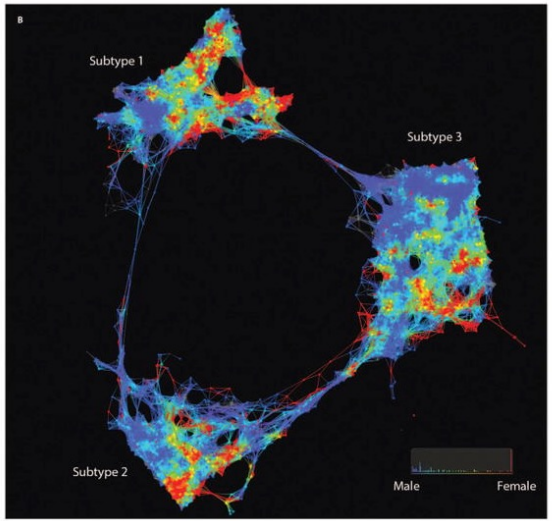
# 3.1 Selection of patient populations for investigation

Decrease sample size

Identify patterns in patient features

# 3.1 Selection of patient populations for investigation

*Electronic health record (EHR) and genetic data identified three different subtypes of type 2 diabetes*[1]




*[1]:* Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med. 2015;7(311):311ra174

# 3.1 Selection of patient populations for investigation

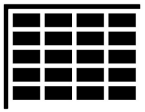
PITFALLS

- Lack negative data
- Limited drug usage
- Missing subgroups

# 3.2 Participant monitoring

1. IDENTIFICATION

NLP

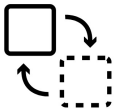


2. DATA ANALYSIS

3. DECREASE STUDY BURDENS

# 4.
# Data collection and management

# 4.1 Wearable and other smart devices



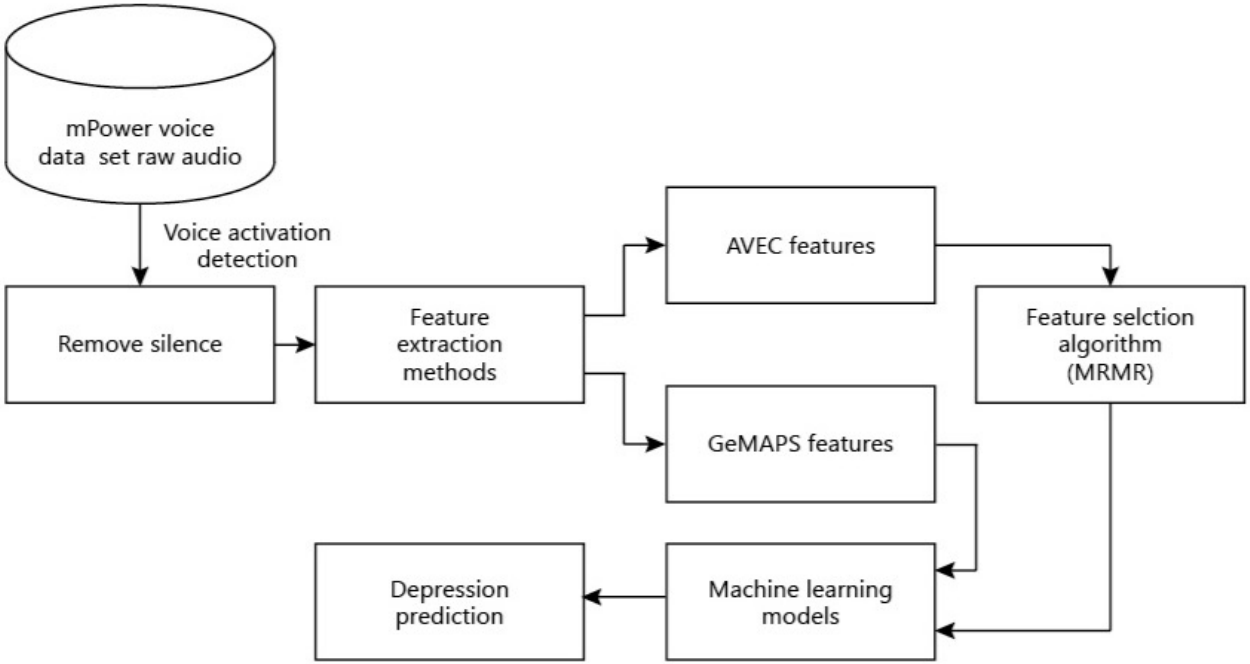
Supplement or even replace study visits

Large and complex data

Ethical and privacy

# 4.1 Wearable and other smart devices

## *Depression Screening from Voice Samples of Patients Affected by Parkinson's Disease*[1]




*[1]:* Ozkanca Y, Ozturk MG, Ekmekci MN, Atkins DC, Demiroglu C, Ghomi RH. Depression screening from voice samples of patients affected by Parkinson's disease. Digit Biomark. 2019;3(2):72–82. https://doi.org/10.1159/000500354.

# 4.2 Data collection and missing data

1. Automate

Case report forms

2. Impute

Covariate values

3. Average

Distribution

# 5.

# Barriers to the integration of ML techniques in clinical research

# 5.1 Operational barriers

Assemble a group

Building models

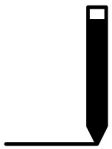
Algorithm development and validation

# 5.2 Philosophical barriers

- Explainability
  - Attention scores
- Trustworthiness
  - Clinical medicine that is not well understood continue to be used

# 6.

# Conclusion



1. Not Evaluated in peer-reviewed manner

2. Distort clinical reality

3. Bias (Ethical, socioeconomic ...)

4. Preclinical rather than clinical trial planning

5. High requirement for data structures and algorithms

*We also hope that ML in clinical research is applied in a fair, ethical, and open manner that is acceptable to all.*

*Questions?*