

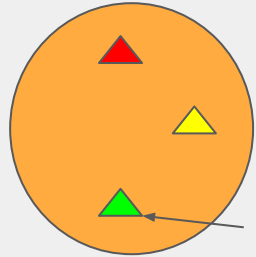
# Sequential Design of Experiments with Unknown Covariates

Harvineet Singh, Rumi Chunara



**NEW YORK UNIVERSITY**

# Optimal Experiment Design



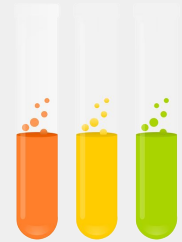
Design space  
e.g. protein structures

Select  
➔



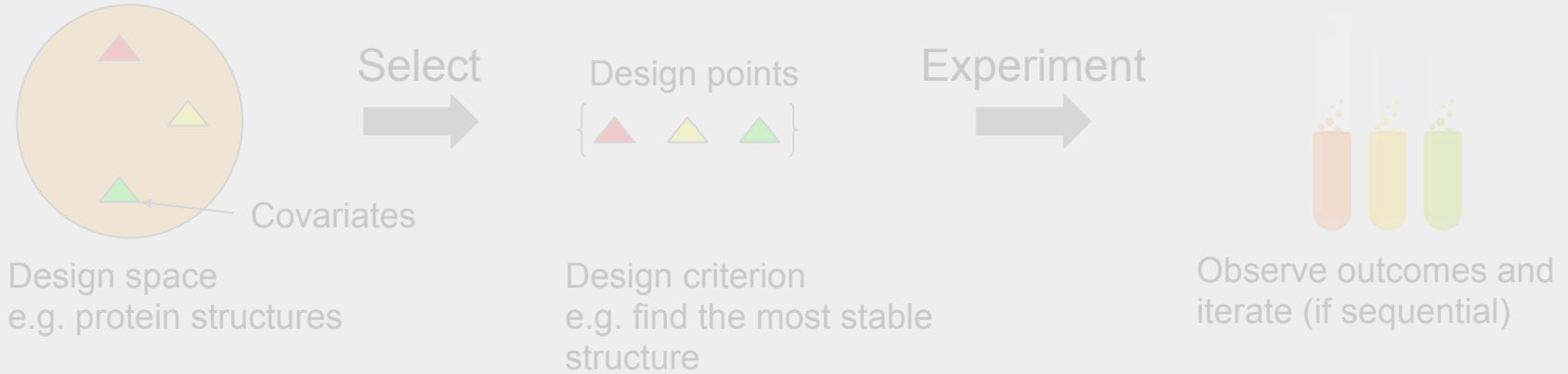
Design criterion  
e.g. find the most stable  
structure

Experiment  
➔

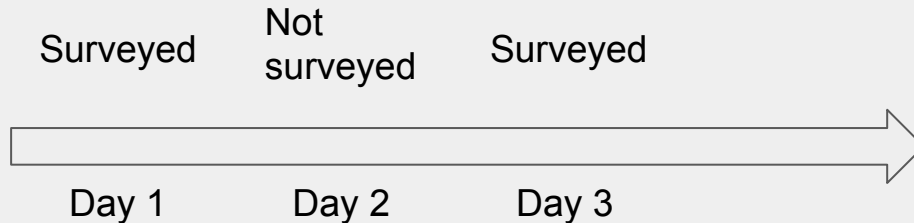
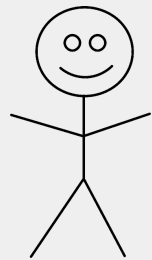


Observe outcomes and  
iterate (if sequential)

# Design for Longitudinal Studies



When covariates are **not observed or not in control** of experimenter



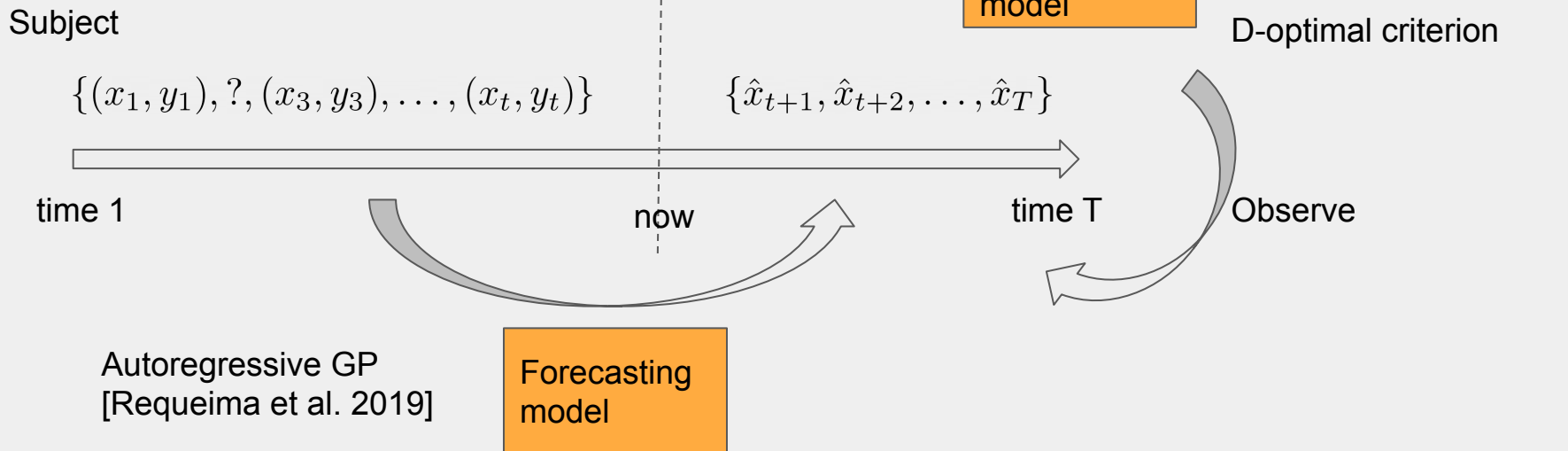
e.g. Behavior assessments  
[Shiffman et al., 2008]

Both covariates and outcomes are not available for selecting design points

# Setup and approach

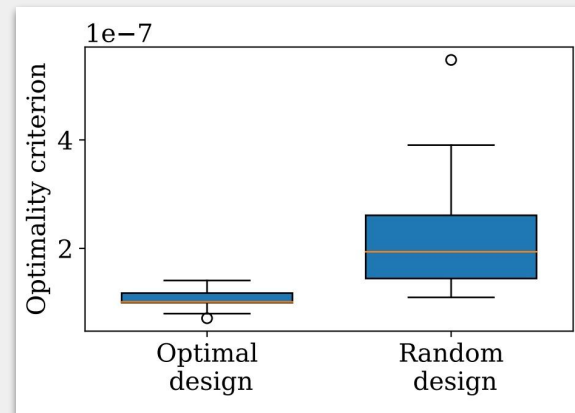
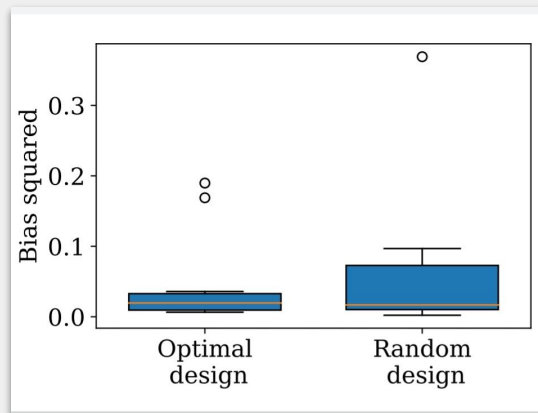
Parameter estimation  
Goal:  
statistical efficiency

$$\text{Observe } (X_{\text{obs}}, Y_{\text{obs}}), \quad Y_{\text{obs}} = f(X_{\text{obs}}; \theta), \quad \hat{\theta} ?$$



# Results - Synthetic data

Compared to uniform random sampling, **similarly low bias and lower variance** for the parameter estimate



## Further work

Characterizing bias due to forecasting model error

Use of forecast uncertainty in selecting designs

More model classes and optimality criteria

# Thanks!

hs3673@nyu.edu

[harvineet.github.io](https://harvineet.github.io)

# References

Saul Shiffman, Arthur A Stone, and Michael R Hufford. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4:1–32, 2008.

Nan M Laird, and James H Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

Valerii Fedorov. Optimal experimental design. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):581–589, 2010.

James Requeima, William Tebbutt, Wessel Bruinsma, and Richard E Turner. The gaussian process autoregressive regression model (gpar). In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1860–1869, 2019.