

# Strategic Hypothesis Testing

*NeurIPS 2025 · Spotlight*



**Yatong Chen\***



**Safwan Hossain\***



**Yiling Chen**

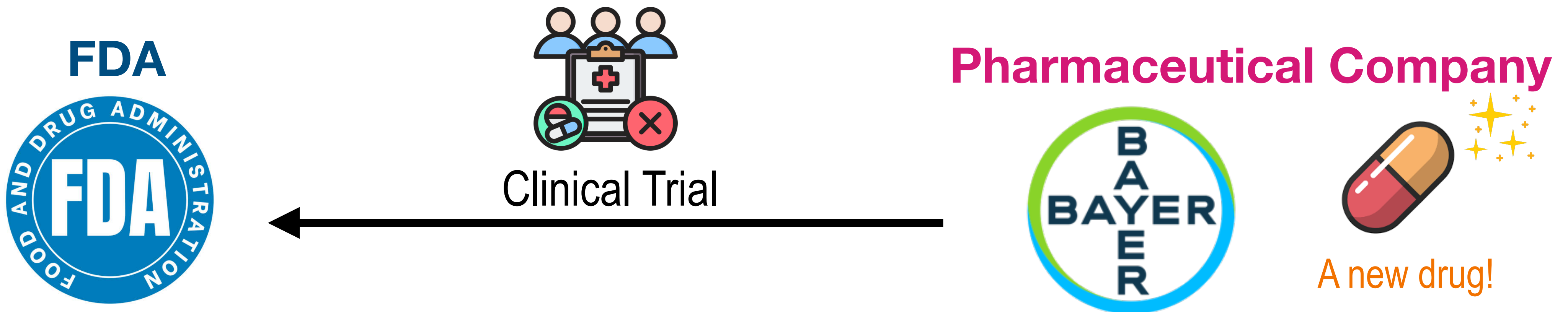
*\*equal contribution*

# Decision-Making Based on Statistical Evidence

- Many real-world decisions rely on (noisy) statistical evidence when ground true is unobservable

# Decision-Making Based on Statistical Evidence

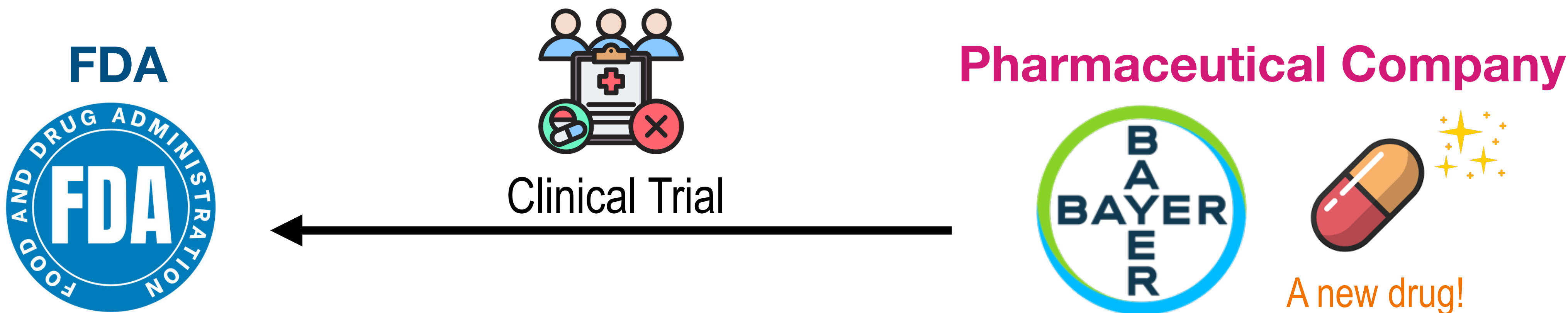
- Many real-world decisions rely on (noisy) statistical evidence when ground true is unobservable
- Example: the FDA's drug approval process



- **FDA's decision** (Approve ✓ or Reject ✗) on the drug is entirely based on clinical trial results

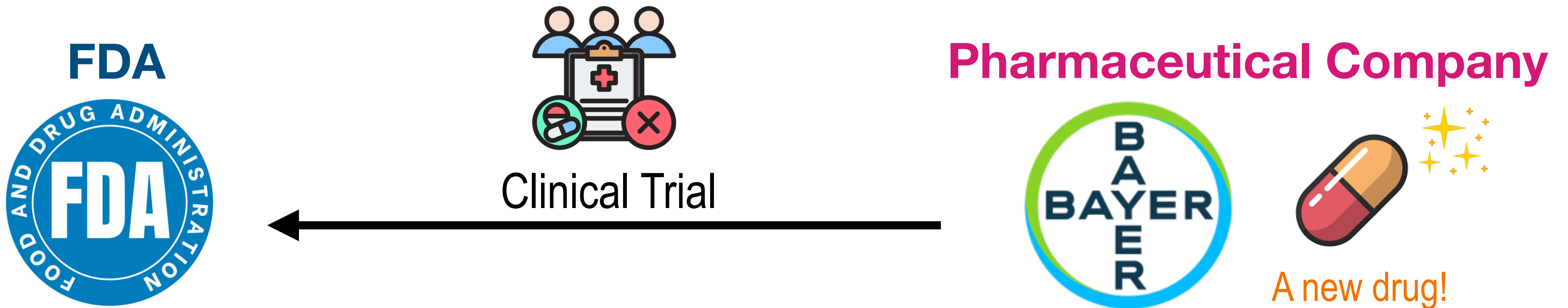


# Decision-Making Based on Statistical Evidence



FDA's Testing Protocol

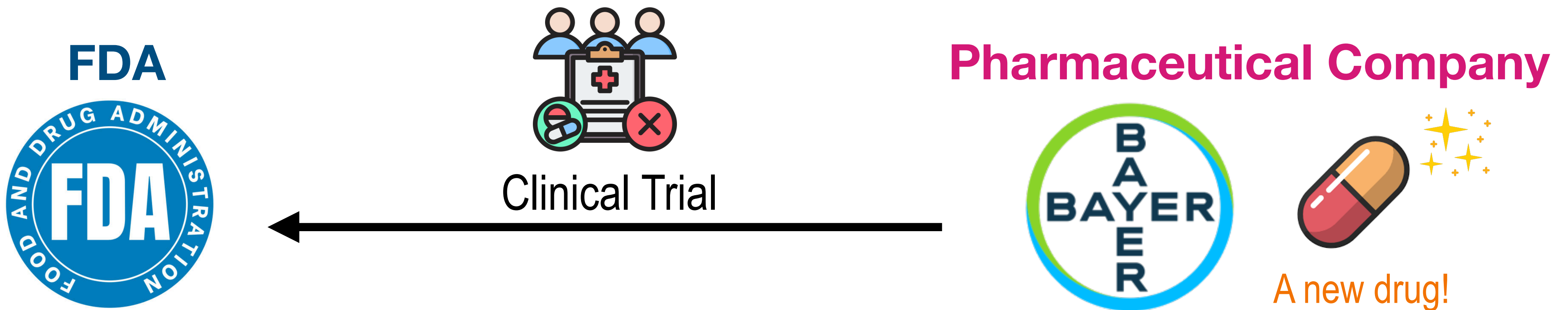
# Decision-Making Based on Statistical Evidence



## FDA's Testing Protocol

**FDA** chooses  
a p-value  
threshold  $\alpha$   
(e.g., 0.05)

# Decision-Making Based on Statistical Evidence



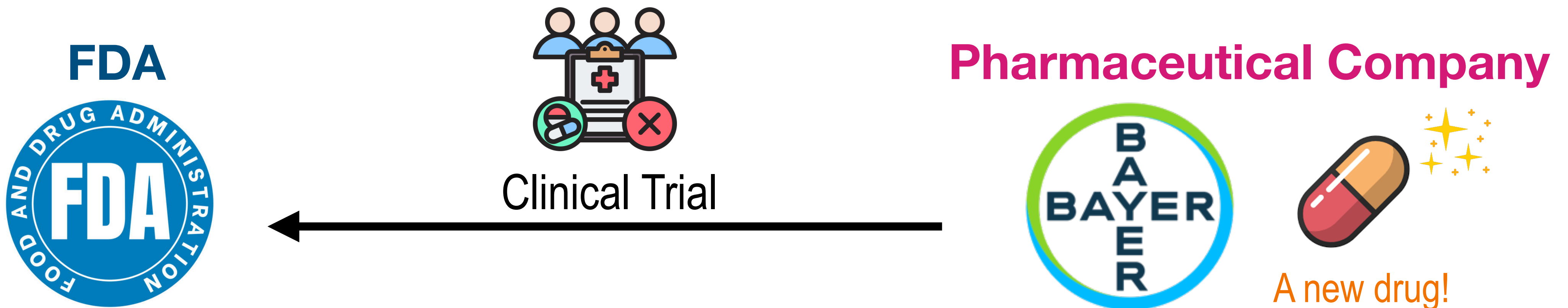
## FDA's Testing Protocol

**FDA** chooses  
a p-value  
threshold  $\alpha$   
(e.g., 0.05)



**Company** runs  
a trial with  $n$   
patients

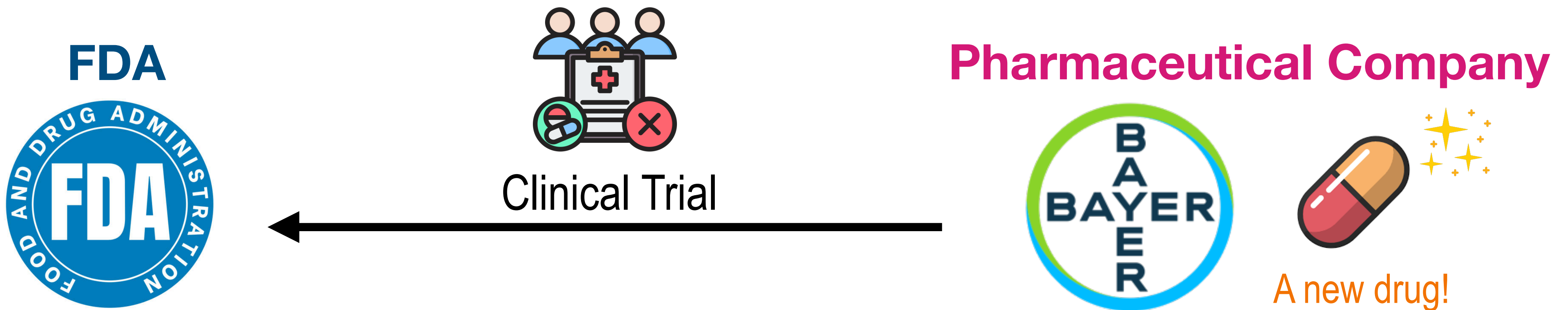
# Decision-Making Based on Statistical Evidence



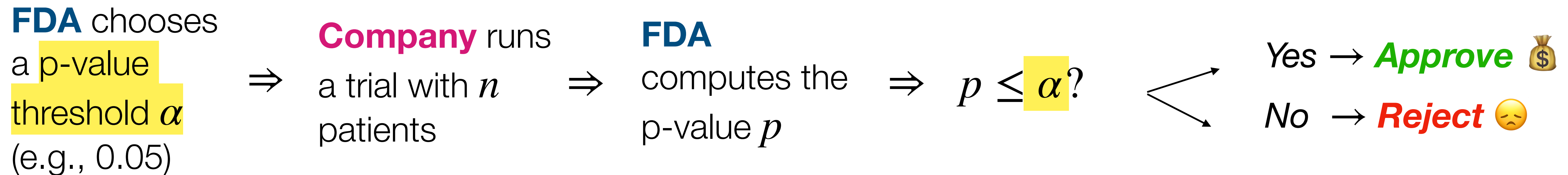
## FDA's Testing Protocol

**FDA** chooses a p-value threshold  $\alpha$  (e.g., 0.05)  $\Rightarrow$  **Company** runs a trial with  $n$  patients  $\Rightarrow$  **FDA** computes the p-value  $p$

# Decision-Making Based on Statistical Evidence

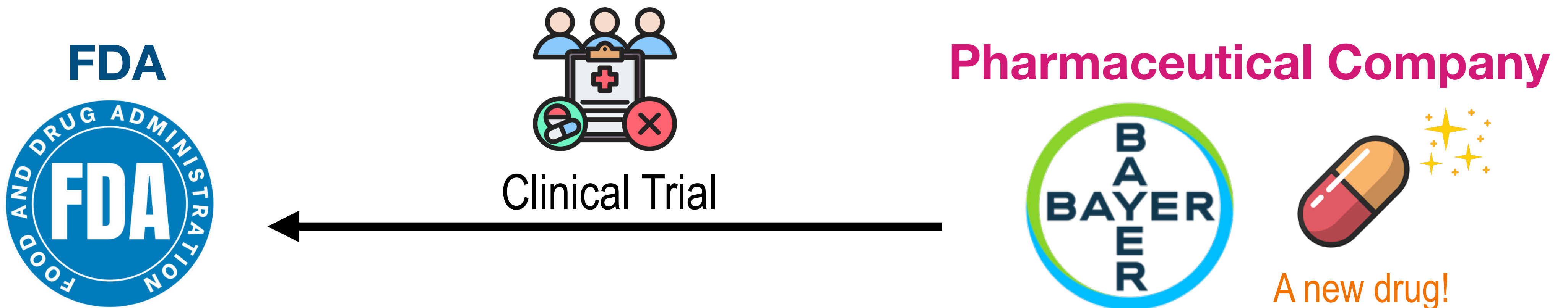


## FDA's Testing Protocol

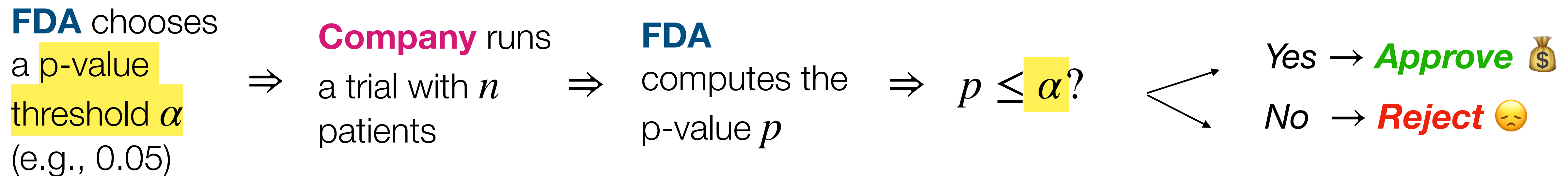




# Decision-Making Based on Statistical Evidence



## FDA's Testing Protocol



This setup induces a **strategic interaction**: each side's decision depends on the other's.

# Stackelberg Game between FDA and the Company

## FDA's Testing Protocol

**FDA** chooses  
a p-value  
threshold  $\alpha$   
(e.g., 0.05)

$\Rightarrow$

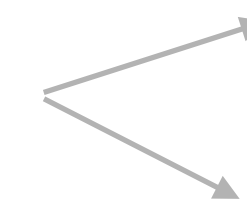
**Company** runs  
a trial with  $n$   
patients

$\Rightarrow$

**FDA**  
computes the  
p-value  $p$

$\Rightarrow$

$p \leq \alpha?$

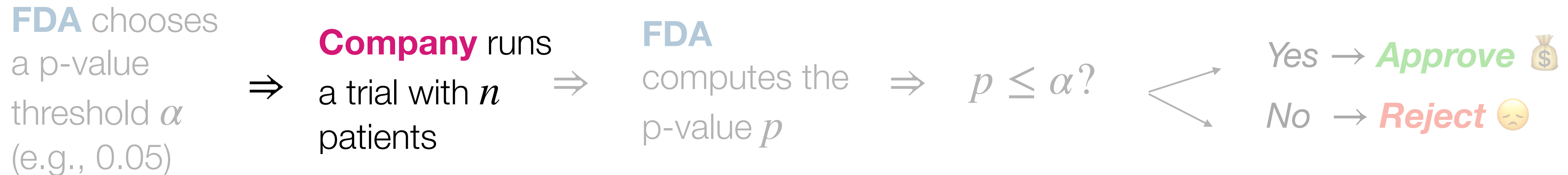


Yes  $\rightarrow$  **Approve** 💰

No  $\rightarrow$  **Reject** 😞



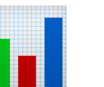
# Drug Company's Decision & Utility

## FDA's Testing Protocol



- **Company** (with effectiveness  $\mu$ ) chooses the size of the clinical trial  $n^*$  to maximize the expected utility

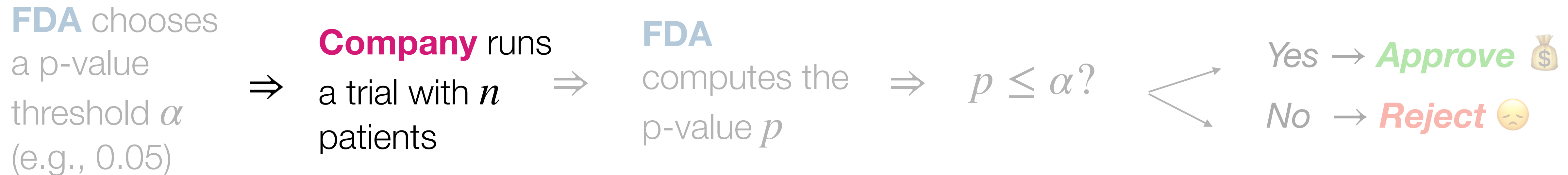
$$u(n; \alpha) = R \cdot \Pr[\text{approve} \mid n, \alpha] - \text{Cost}(n)$$

Revenue once approved   

Cost of the trial (with  $n$  samples)  
  

# Drug Company's Decision & Utility

## FDA's Testing Protocol



- **Company** (with effectiveness  $\mu$ ) chooses the size of the clinical trial  $n^*$  to maximize the expected utility

For any fixed rule  $\alpha$ , trial size  $n \uparrow \Rightarrow$   
 $Pr[\text{approve} | n, \alpha] \uparrow$  but  $Cost(n) \uparrow$

$$u(n; \alpha) = R \cdot \Pr[\text{approve} | n, \alpha] - \text{Cost}(n)$$

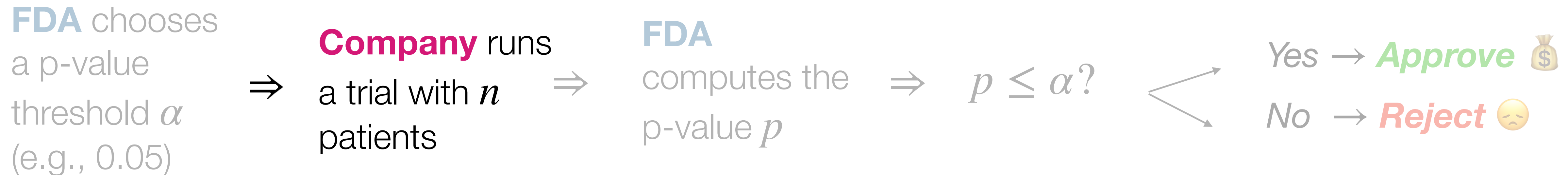
Revenue once approved  
📈💰📊

Cost of the trial  
(with  $n$  samples)  
💵💵💵



# Drug Company's Decision & Utility

## FDA's Testing Protocol



- **Company** (with effectiveness  $\mu$ ) chooses the size of the clinical trial  $n^*$  to maximize the expected utility

For any fixed rule  $\alpha$ , trial size  $n \uparrow \Rightarrow$   
 $Pr[\text{approve} | n, \alpha] \uparrow$  but  $Cost(n) \uparrow$

$$u(n; \alpha) = R \cdot \Pr[\text{approve} | n, \alpha] - \text{Cost}(n)$$

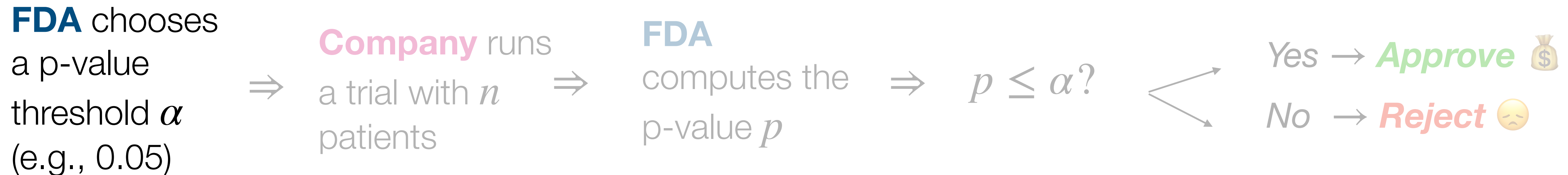
Revenue once approved  
📈💰📊

Cost of the trial  
(with  $n$  samples)  
💵💵💵

**Company** will do strategic self-selection:  
If  $u(n^*; \alpha) < 0$ , agent will not participate

# FDA's Decision & Utility

## FDA's Testing Protocol



- **Company** (with effectiveness  $\mu$ ) chooses the size of the clinical trial  $n^*$  to maximize the expected utility

$$u(n; \alpha) = R \cdot \Pr[\text{approve} \mid n, \alpha] - \text{Cost}(n)$$

Revenue once approved 

Cost of the trial (with  $n$  samples) 

- **The FDA**, knowing the company will do a **strategic self-selection**, chooses  $\alpha^*$  to minimize total errors

$$\ell(\alpha) = \mathbb{E}_Q[\lambda_1 \text{FP} + \lambda_2 \text{FN}]$$

Approve inefficient drug

Disapprove effective drug/  
effective drug not participate

# Result & Implications

- We provide characterizations of **optimal behavior** of both **the principal** (selection criteria  $\alpha^*$ ) and **the agent** (optimal size of the trial  $n^*$ )

# Result & Implications

- We provide characterizations of **optimal behavior** of both **the principal** (selection criteria  $\alpha^*$ ) and **the agent** (optimal size of the trial  $n^*$ )
- We characterize how different loss components (FP, FN) change as we vary  $\alpha$ 
  - They are still monotonic in  $\alpha$  even when agents are strategically!



# Result & Implications

- We provide characterizations of **optimal behavior** of both **the principal** (selection criteria  $\alpha^*$ ) and **the agent** (optimal size of the trial  $n^*$ )
- We characterize how different loss components (FP, FN) change as we vary  $\alpha$ 
  - They are still monotonic in  $\alpha$  even when agents are strategically!
- Our results offer insights for **policymakers/regulators**

# Result & Implications

- We provide characterizations of **optimal behavior** of both **the principal** (selection criteria  $\alpha^*$ ) and **the agent** (optimal size of the trial  $n^*$ )
- We characterize how different loss components (FP, FN) change as we vary  $\alpha$ 
  - They are still monotonic in  $\alpha$  even when agents are strategically!
- Our results offer insights for **policymakers/regulators**

**Statistical decision rules not only determine who passes the test, but also who chooses to enter in the first place!**

# Thank you!

