Predictive

~~Big Data~~ Analytics (ISOM 5270)

by Prof. Fernández-Loría

There won’t be a quiz today.

Announcement:
Assignment #2 will be released later today.

The Plan

~~Fundamentals~~

Week 1: ~~Problem formulation.~~

Week 2: ~~Modeling (Part 1).~~

Week 3: ~~Modeling (Part 2).~~

Week 4: ~~Model Evaluation.~~

(What I think are)

The Big Three

Week 5: Causal analytics.

Week 6: ~~Big (unstructured) data.~~

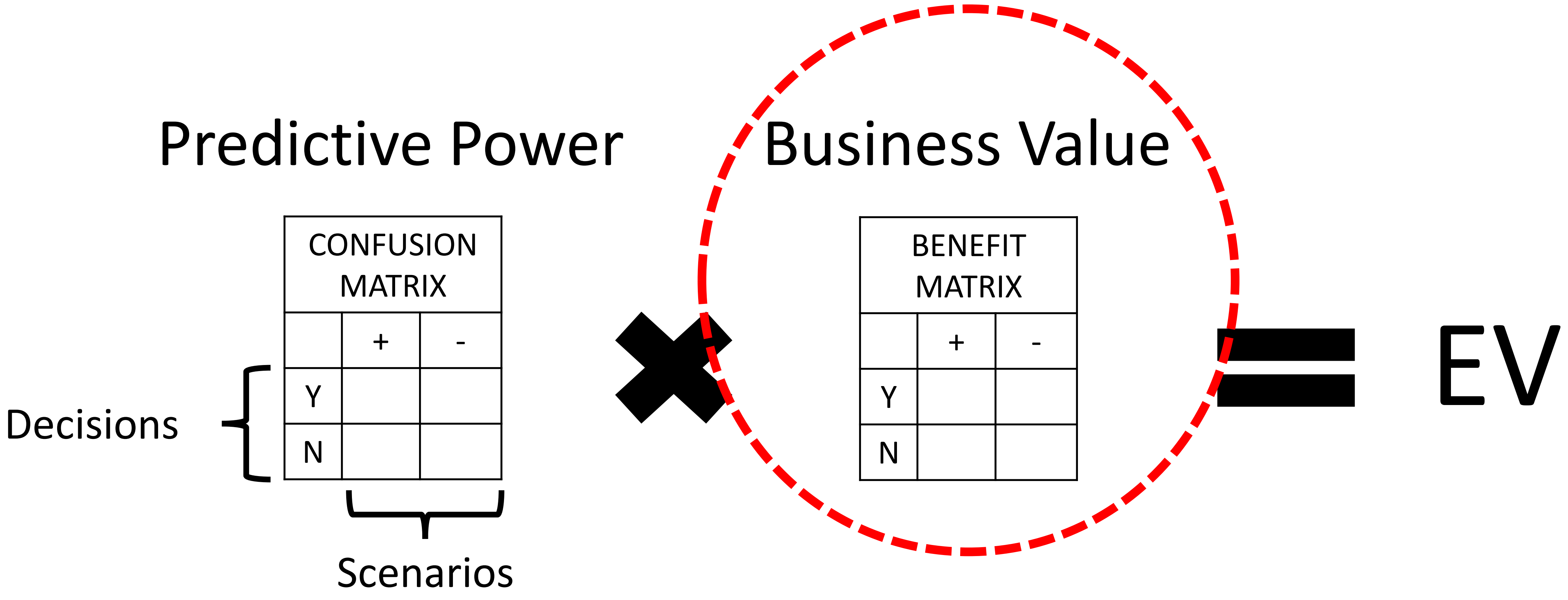
Week 7: Generative AI (e.g., ChatGPT).

Week 8: Final Exam.

Let’s start with the TelCo case study.

Q: Which specific customers should we focus on when offering retention incentives?

Framework #6: Expected Value



Benefit Matrix

	Stay (S)	**Leave (L)**
Target (T)	?	?
Not Target (NT)	?	?

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	Leave (L)
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = ?$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	**Leave (L)**
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = p(S|T) \times (v_T - \$205) - \left(1 - p(S|T)\right) \times \$5$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	**Leave (L)**
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = p(S|T) \times (v_T - \$200) - \$5$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	Leave (L)
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = p(S|T) \times (v_T - \$200) - \$5$$

$$EV_{NT} = ?$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	Leave (L)
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = p(S|T) \times (v_T - \$200) - \$5$$

$$EV_{NT} = p(S|NT) \times v_{NT} - \left(1 - p(S|NT)\right) \times \$0$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	Leave (L)
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = p(S|T) \times (v_T - \$200) - \$5$$

$$EV_{NT} = p(S|NT) \times v_{NT}$$

We should target if doing so is profitable.

$$EV_T > EV_{NT}$$

	Stay (S)	Leave (L)
Target (T)	$v_T - \$205$	$-\$5$
Not Target (NT)	v_{NT}	$\$0$

$$EV_T = p(S|T) \times (v_T - \$200) - \$5$$
$$EV_{NT} = p(S|NT) \times v_{NT}$$

Questions?

We should target if doing so is profitable.

$$p(S|T) \times (v_T - \$200) - \$5 > p(S|NT) \times v_{NT}$$

We should target if doing so is profitable.

$$p(S|T) \times v_T - p(S|NT) \times v_{NT} > p(S|T) \times \$200 + \$5$$

Benefit

Cost

$$p(S|T) \times v_T - p(S|NT) \times v_{NT} > p(S|T) \times \$200 + \$5$$

Is there data to calculate or estimate this quantity?

In the absence of adequate data…
…we could make some assumptions.

If:

$$v_T = v_{NT} \qquad p(S|T) = 1$$

Then:

$$p(S|T) \times v_T - p(S|NT) \times v_{NT} > p(S|T) \times \$200 + \$5$$

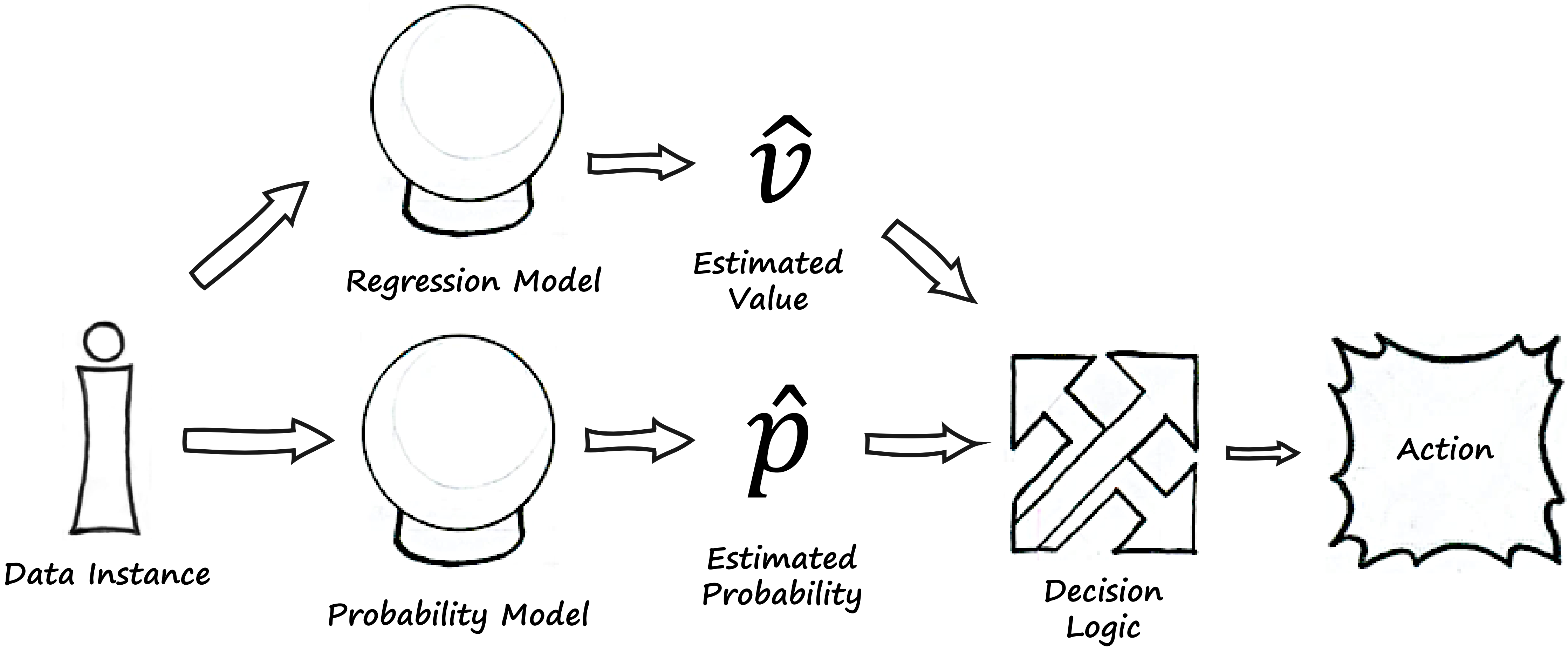
Becomes:

$$p(Leave|NT) \times v_{NT} > \$205$$

Questions?

Key Takeaway

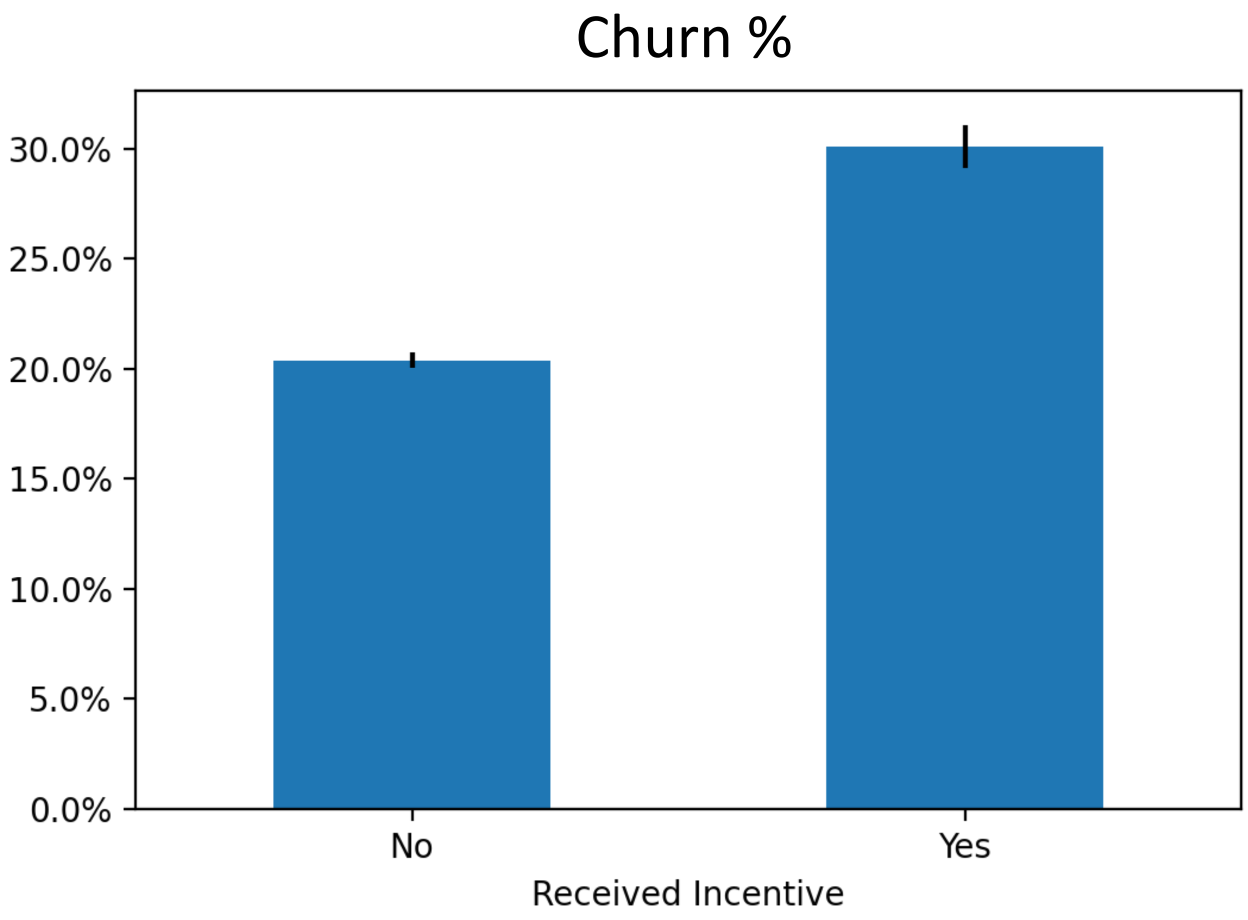
Our decision logic can incorporate multiple quantities (possibly from multiple models).



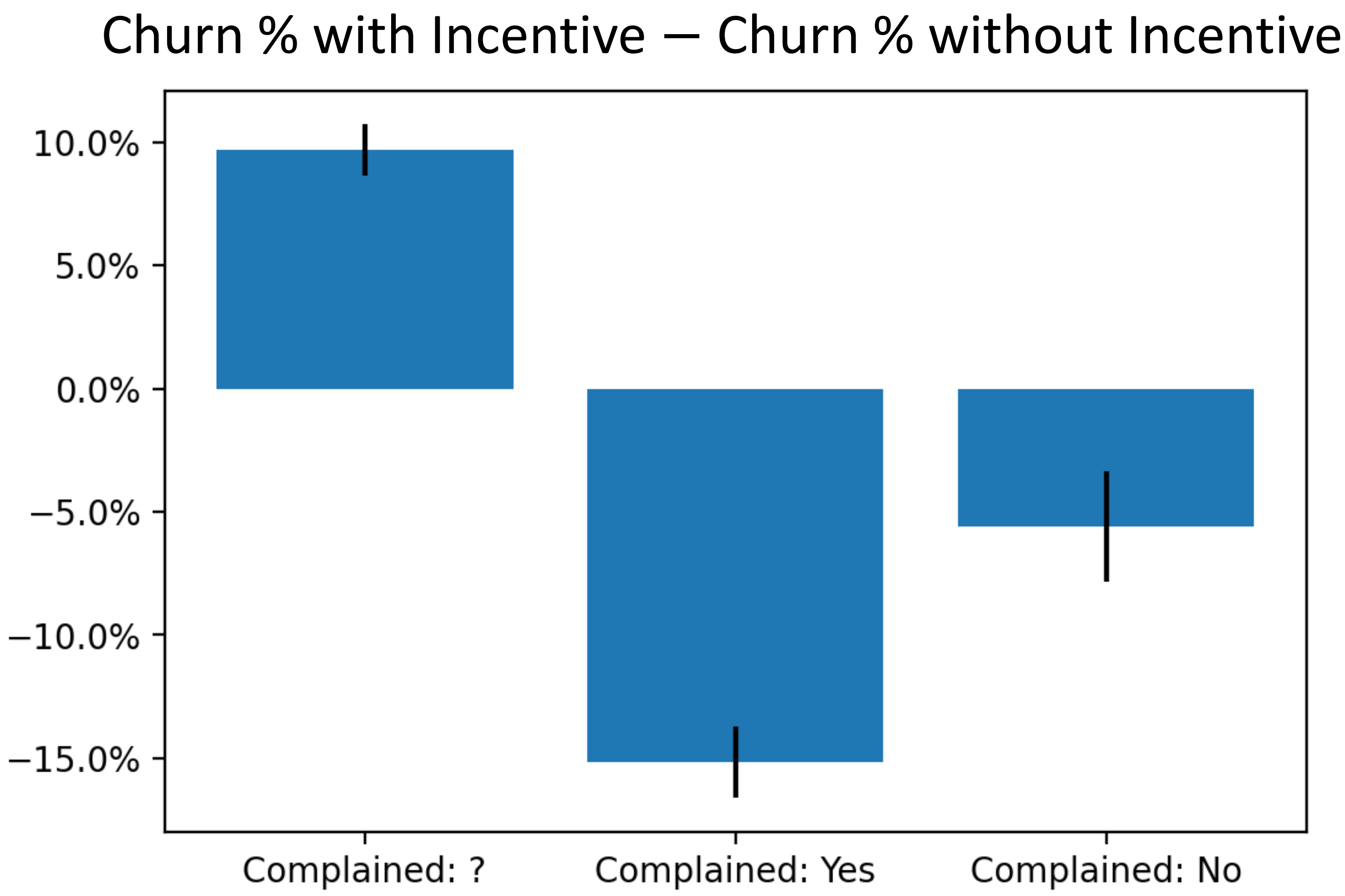
What other alternatives, besides making assumptions, do we have?

What key factors should we consider when retrieving data from previous marketing campaigns?

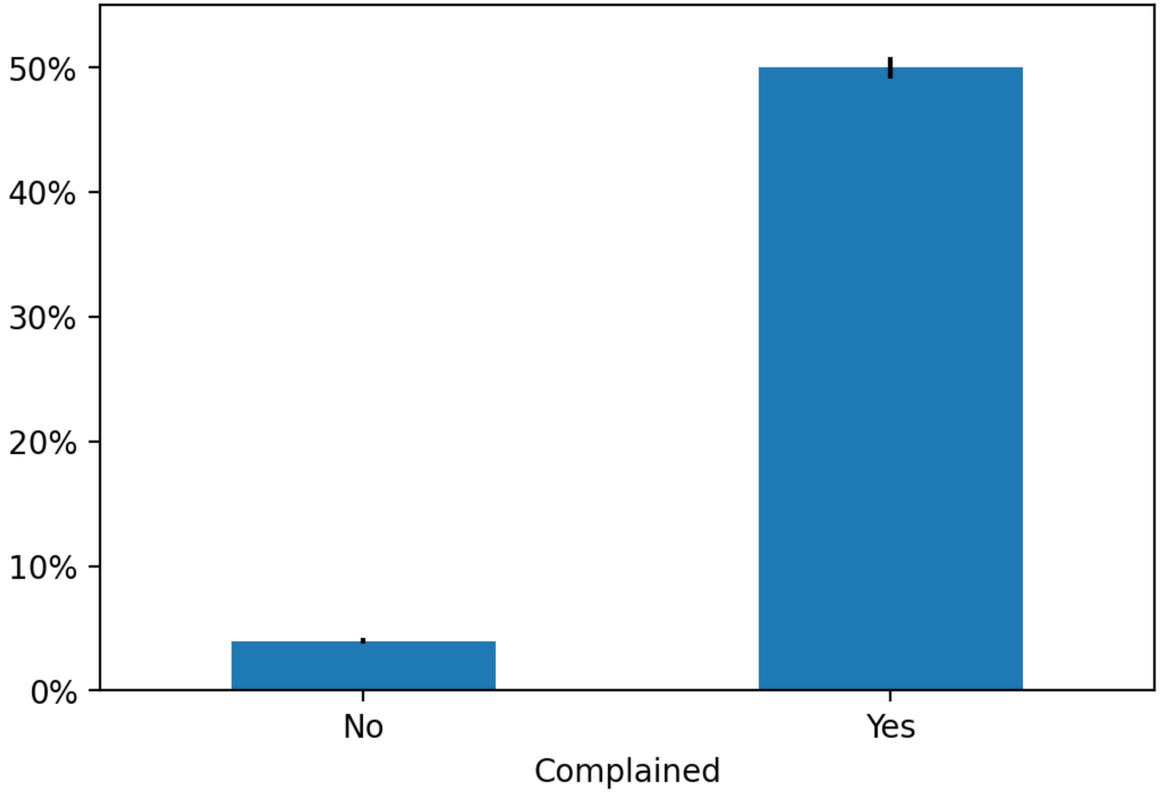
How should we interpret this chart from a past campaign?



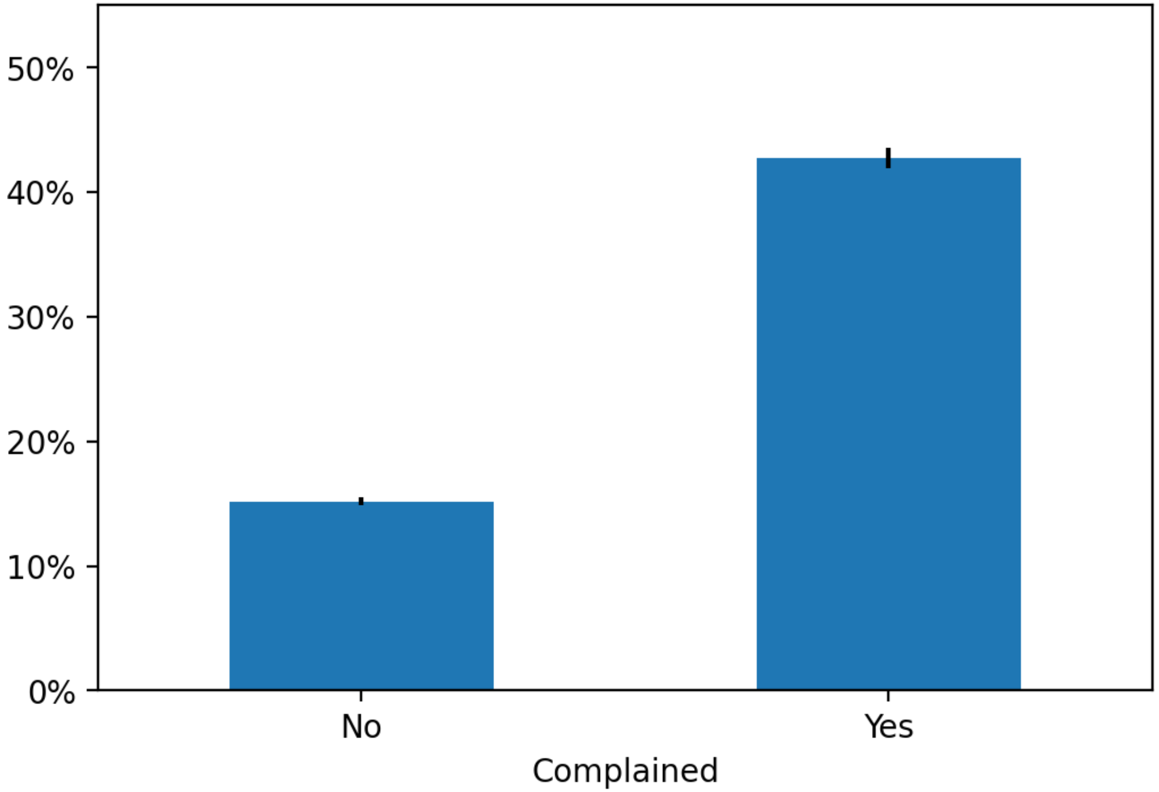
Let’s zoom in. How is this result possible?



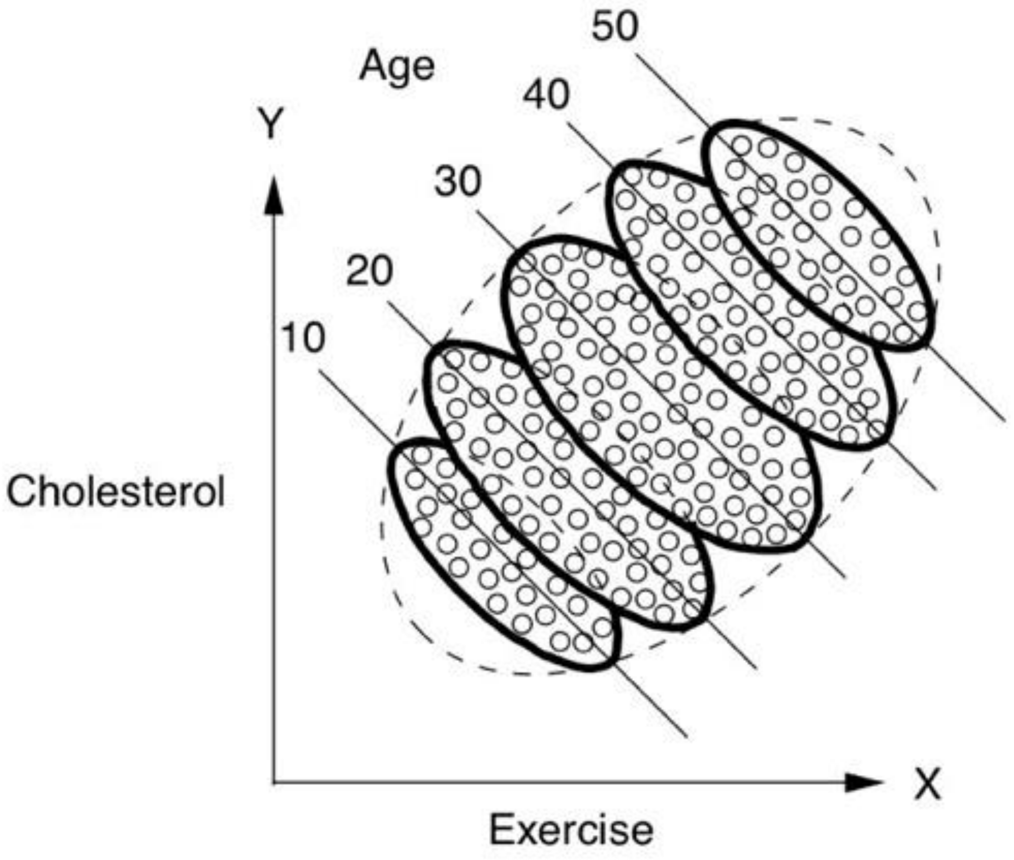
% Receiving Incentive

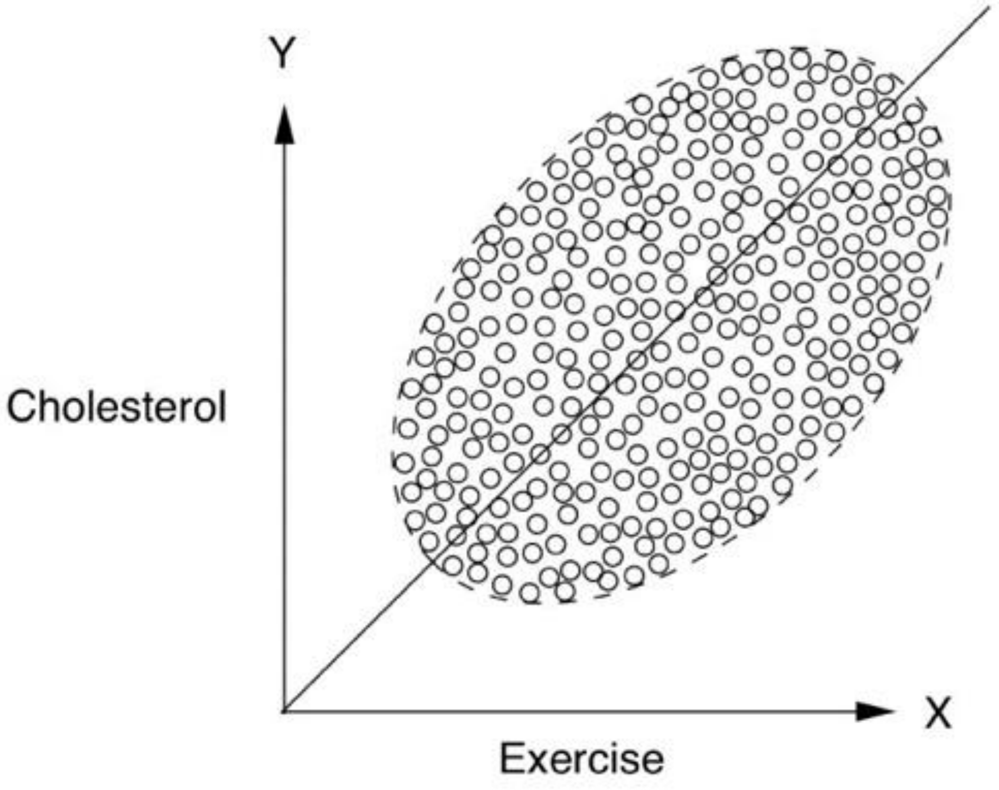


Churn %

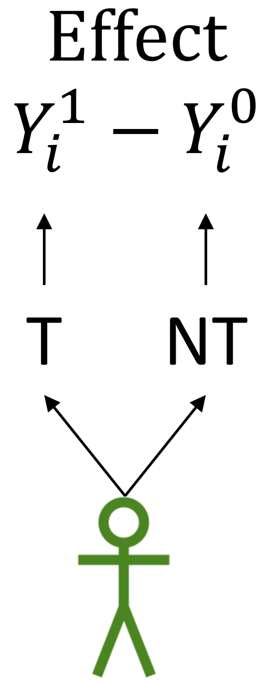


Simpson’s Paradox





To infer causation from correlation,
we need an **apples-to-apples comparison**.



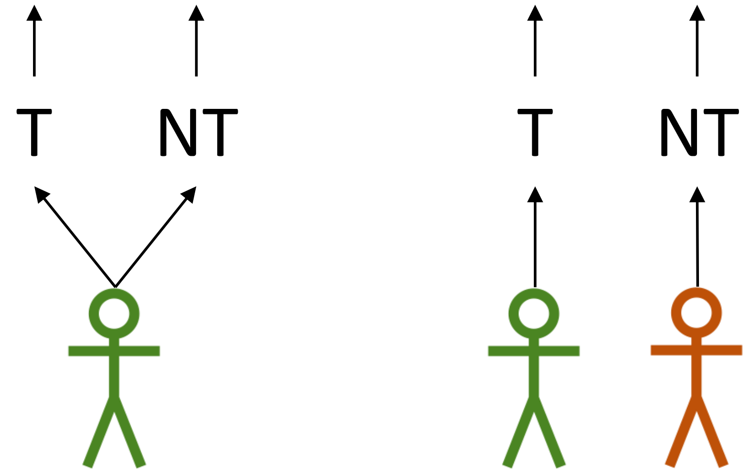
A causal effect is the difference between two **<u>potential outcomes</u>** (alternative realities).

Effect

$Y_i^1 - Y_i^0$

Estimate

$Y_i^1 - Y_j^0$



But we only observe
<u>one</u> potential outcome for each person.

To estimate effects, we compare individuals exposed to different conditions.

<u>Key Assumptions</u>:

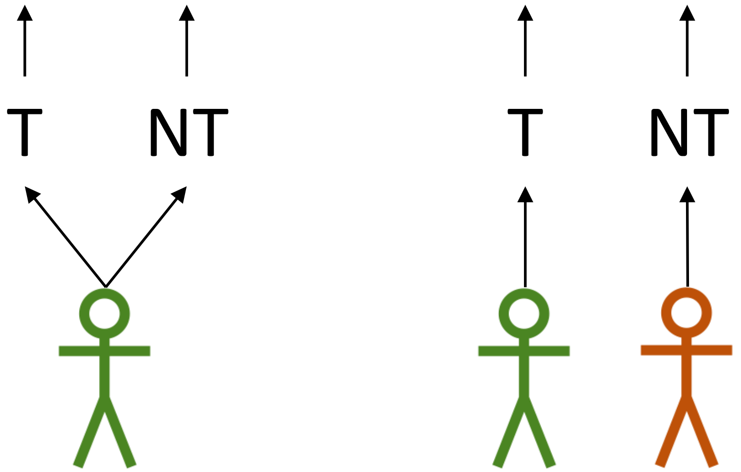
1. **Experimentation**: 1s & 0s

Effect

$$Y_i^1 - Y_i^0$$

Estimate

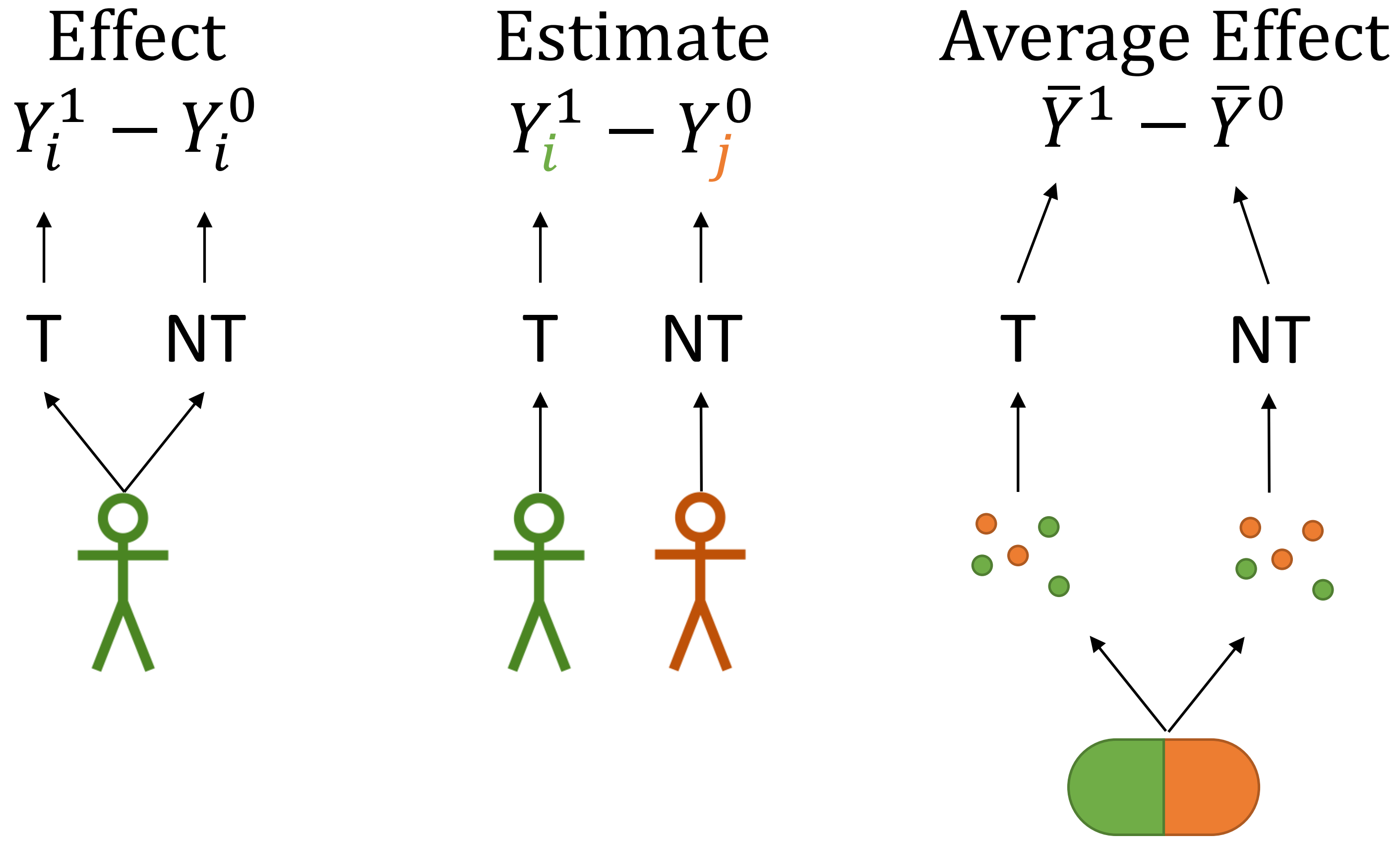
$$Y_i^1 - Y_j^0$$



<u>Key Assumptions</u>:

1. **Experimentation**: 1s & 0s

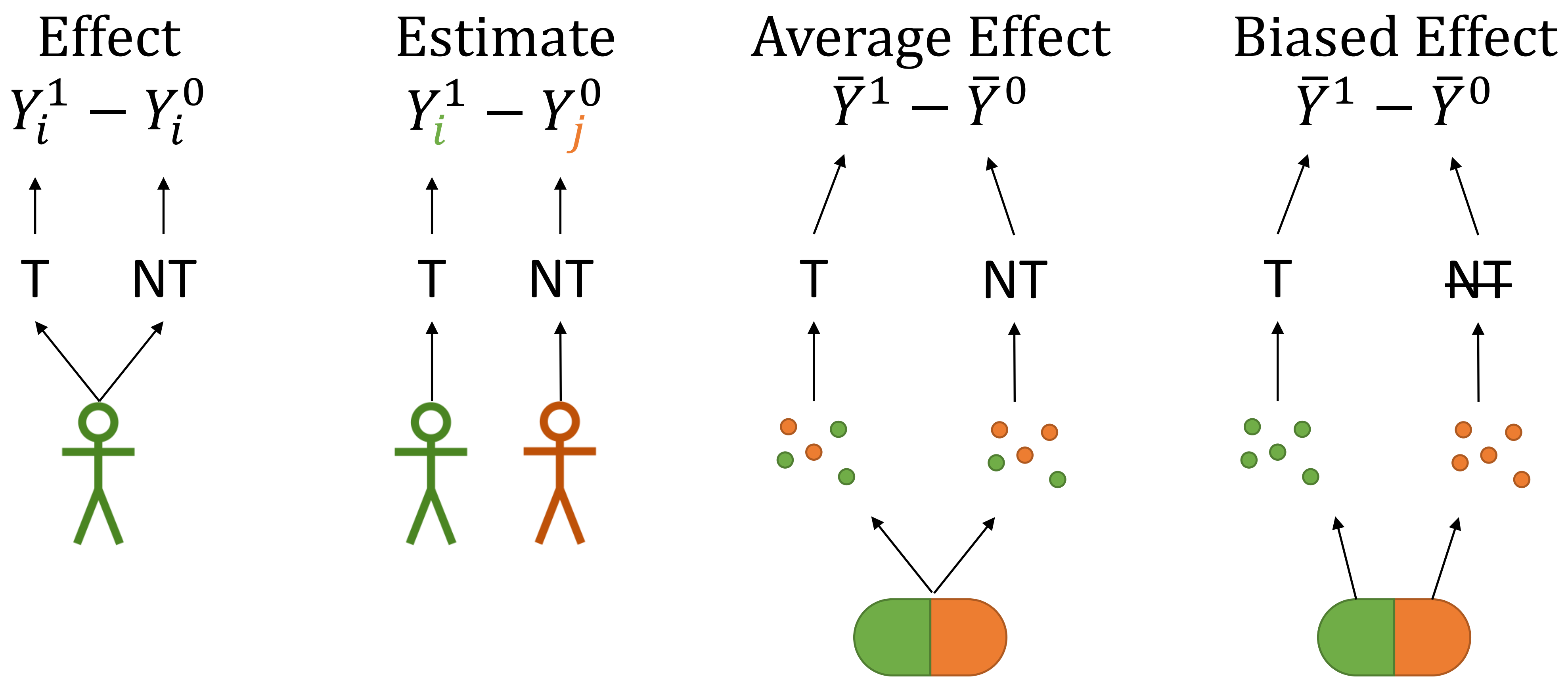
Any concern?



<u>Key Assumptions</u>:

1. **Experimentation**: 1s & 0s

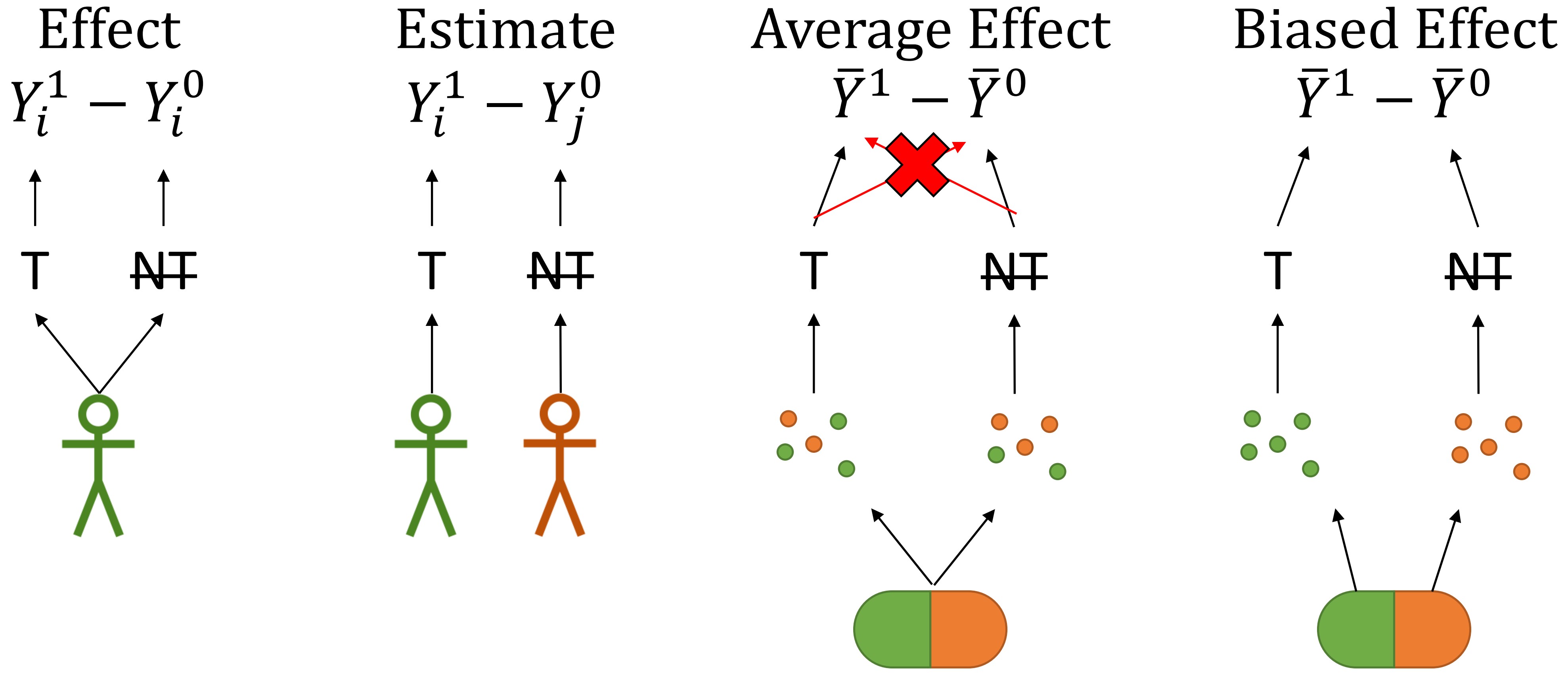
This is where Statistics comes in.



<u>Key Assumptions</u>:

1. **Experimentation**: 1s & 0s
2. **No confounding**: Comparable groups

But we must be careful!



Super careful!

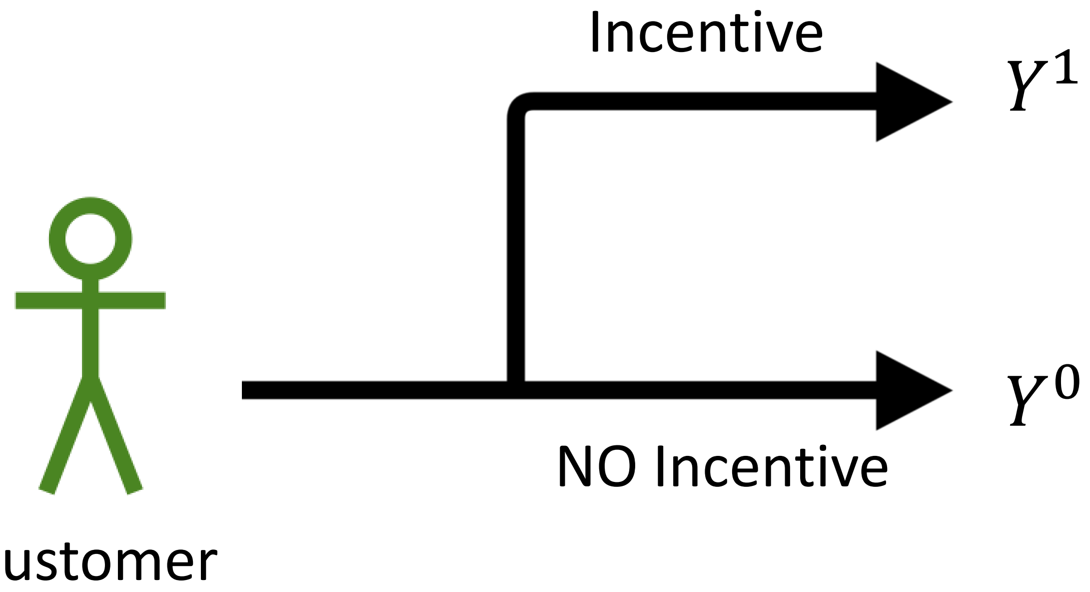
<u>Key Assumptions</u>:

1. **Experimentation**: 1s & 0s
2. **No confounding**: Comparable groups
3. **No contamination**: Each treatment affects one individual.

Questions?

Key Takeaway

A causal effect is the difference between two **potential** outcomes.



$$\text{Effect} = Y^1 - Y^0$$

Key Takeaway

We can't observe both potential outcomes. So:

We estimate effects by comparing people exposed to different conditions.

Give Incentive

Do NOT Give Incentive

Key Takeaway

Beware of treatments that could affect multiple people.

No reminder

Spending **drops**.

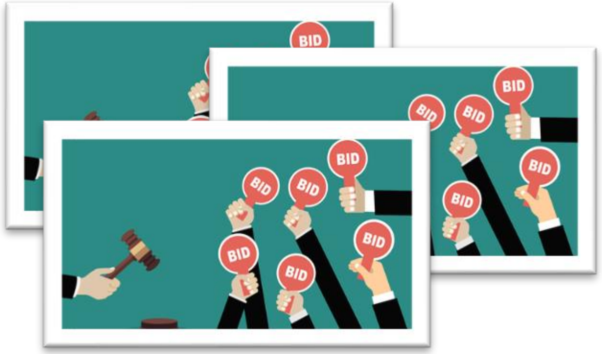
AUCTION

Reminder

Spending increases.

Key Takeaway

Beware of treatments that could affect multiple people.



Spending at the **auction level** with no reminder.

Fix:

ebay AUCTION

vs.

Spending at the **auction level** with reminder.

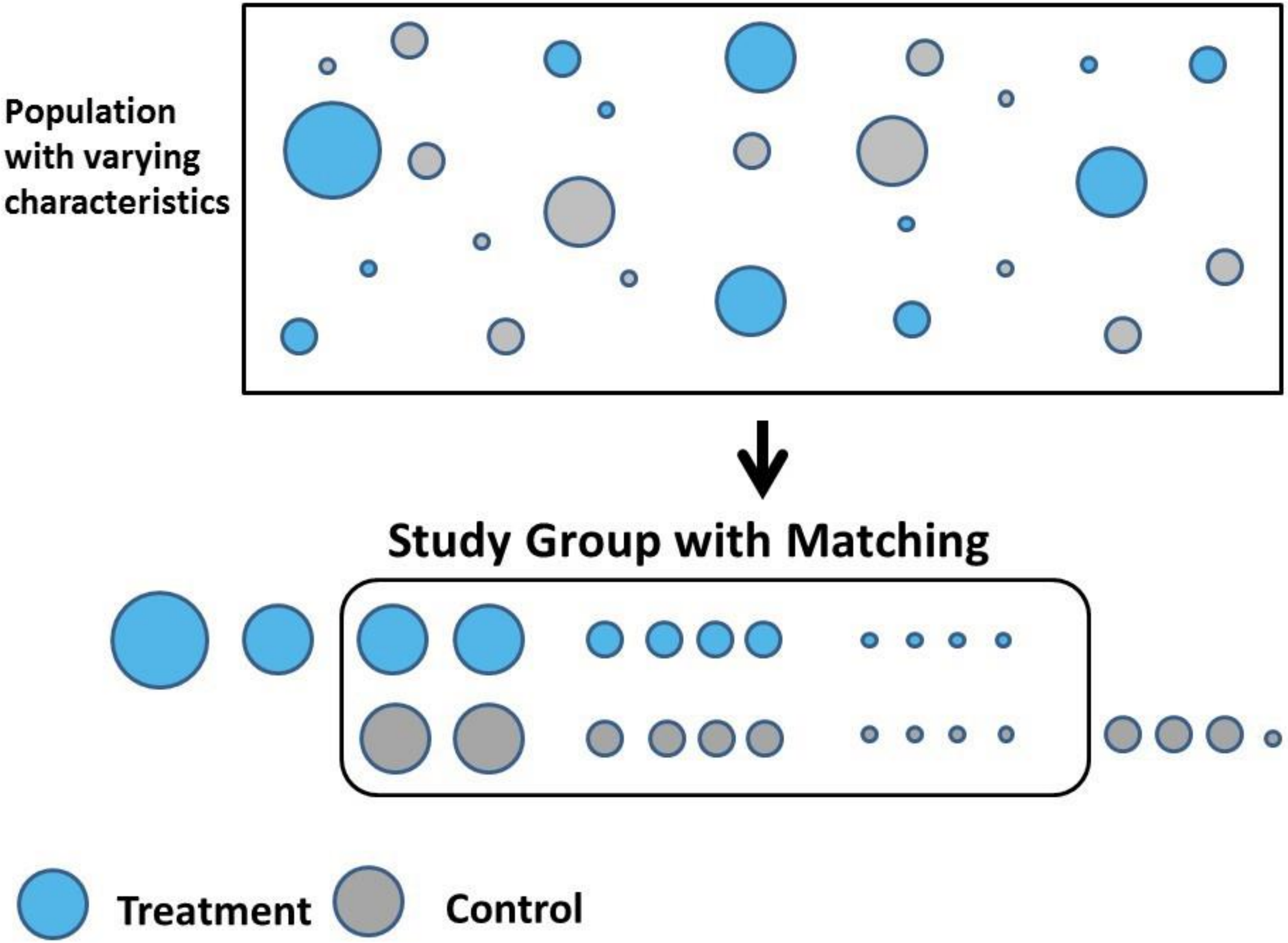
Key Takeaway

To infer causality, we must make apples-to-apples comparisons.

Churn rate with incentive − Churn rate without incentive ≠ Incentive Effect

Churn rate with incentive − Churn rate without incentive = Incentive Effect

How can we make
apples-to-apples comparison?

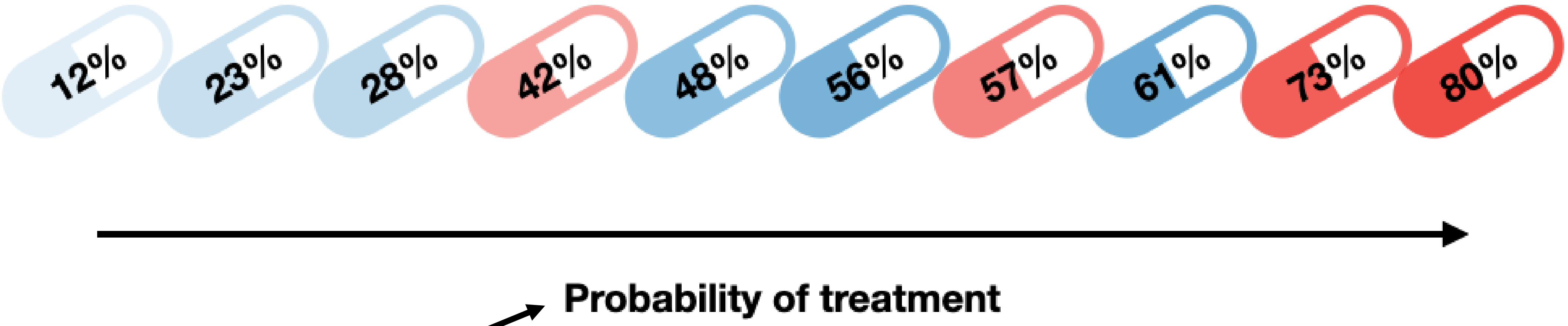

Population
with varying
characteristics
Study Group with Matching
Treatment
Control

But how do we match these?

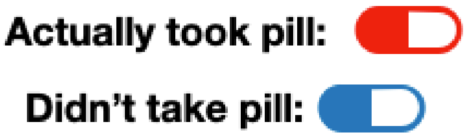
Control

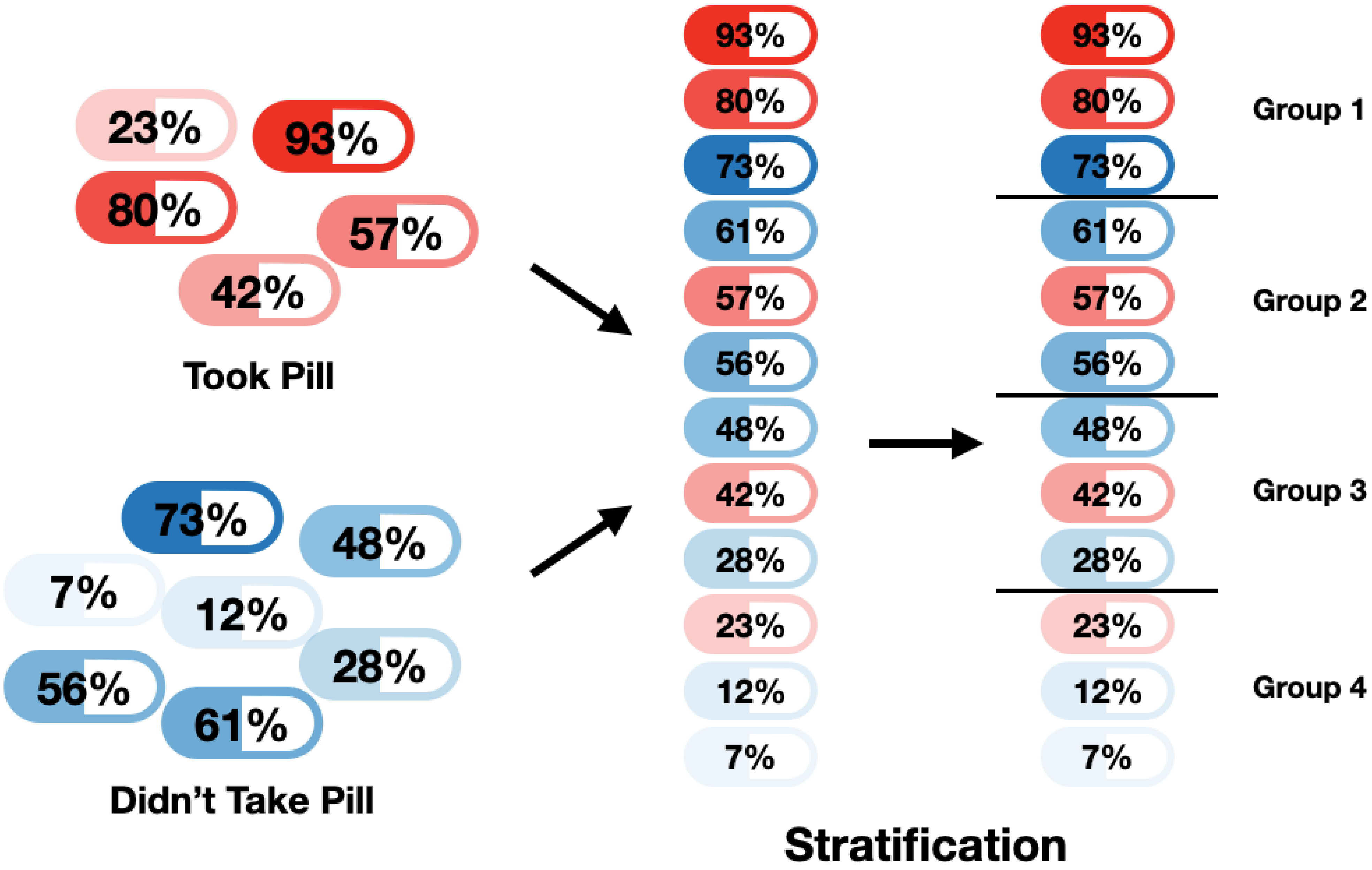
Treatment

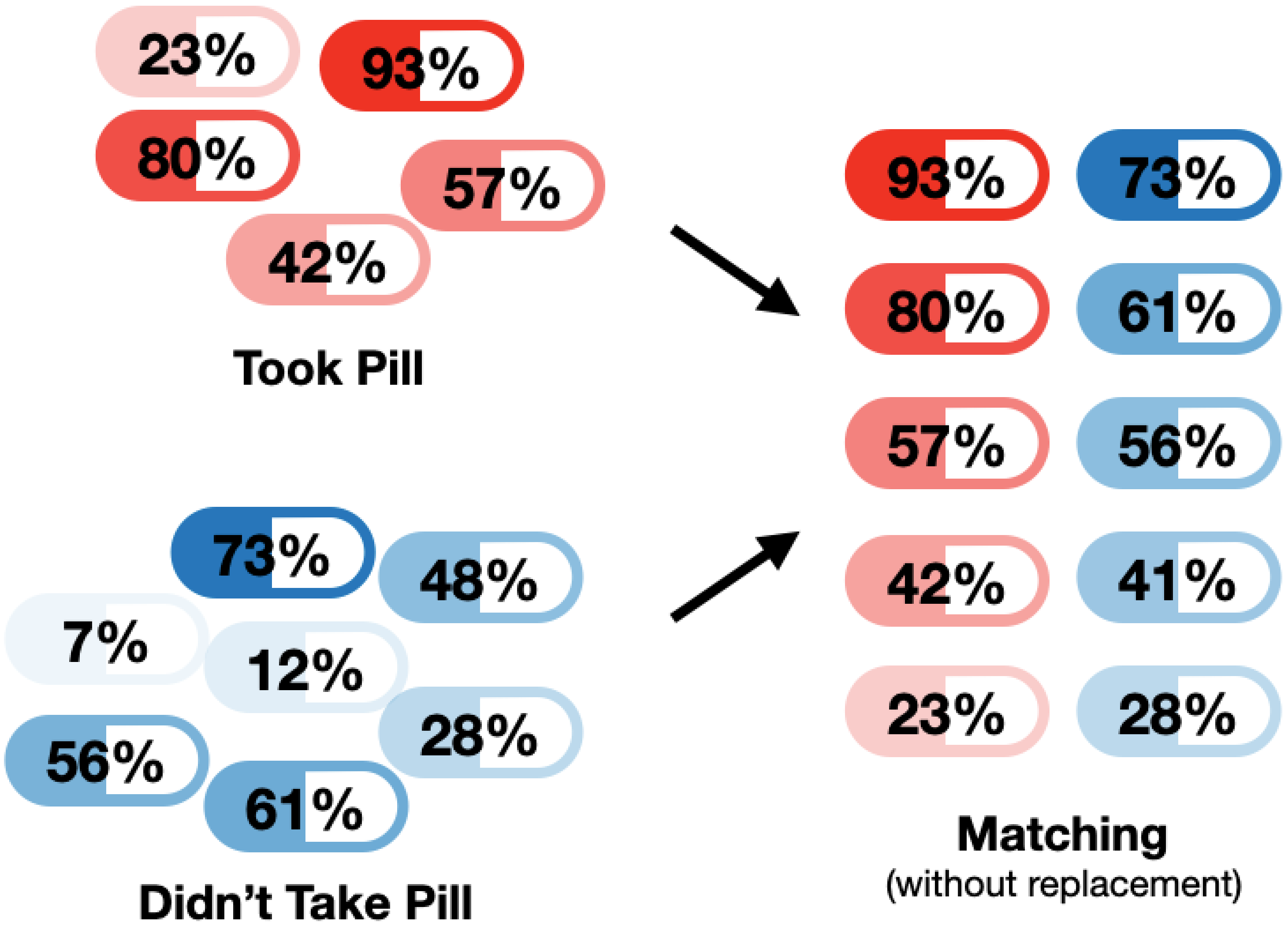
Propensity Score

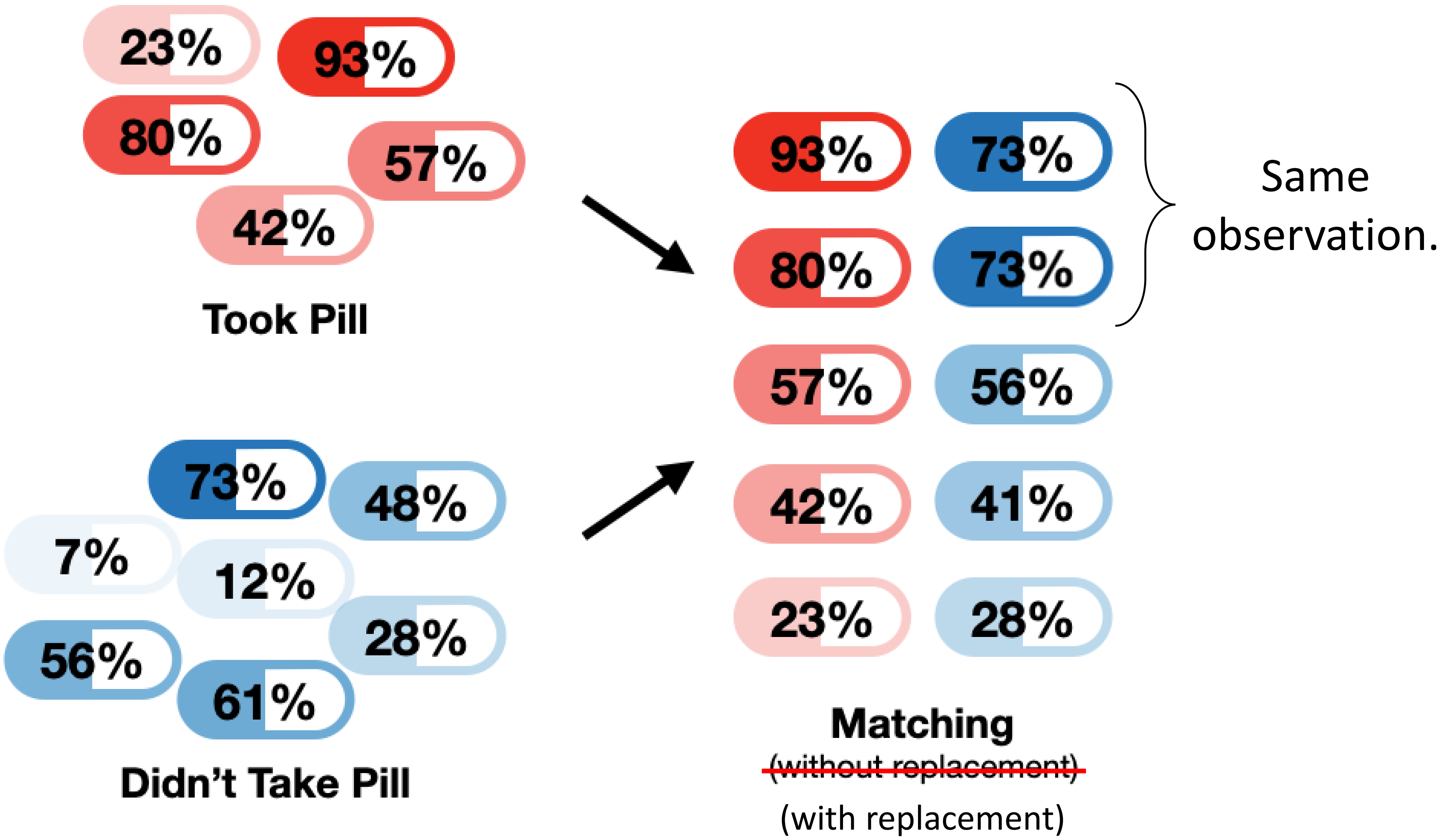


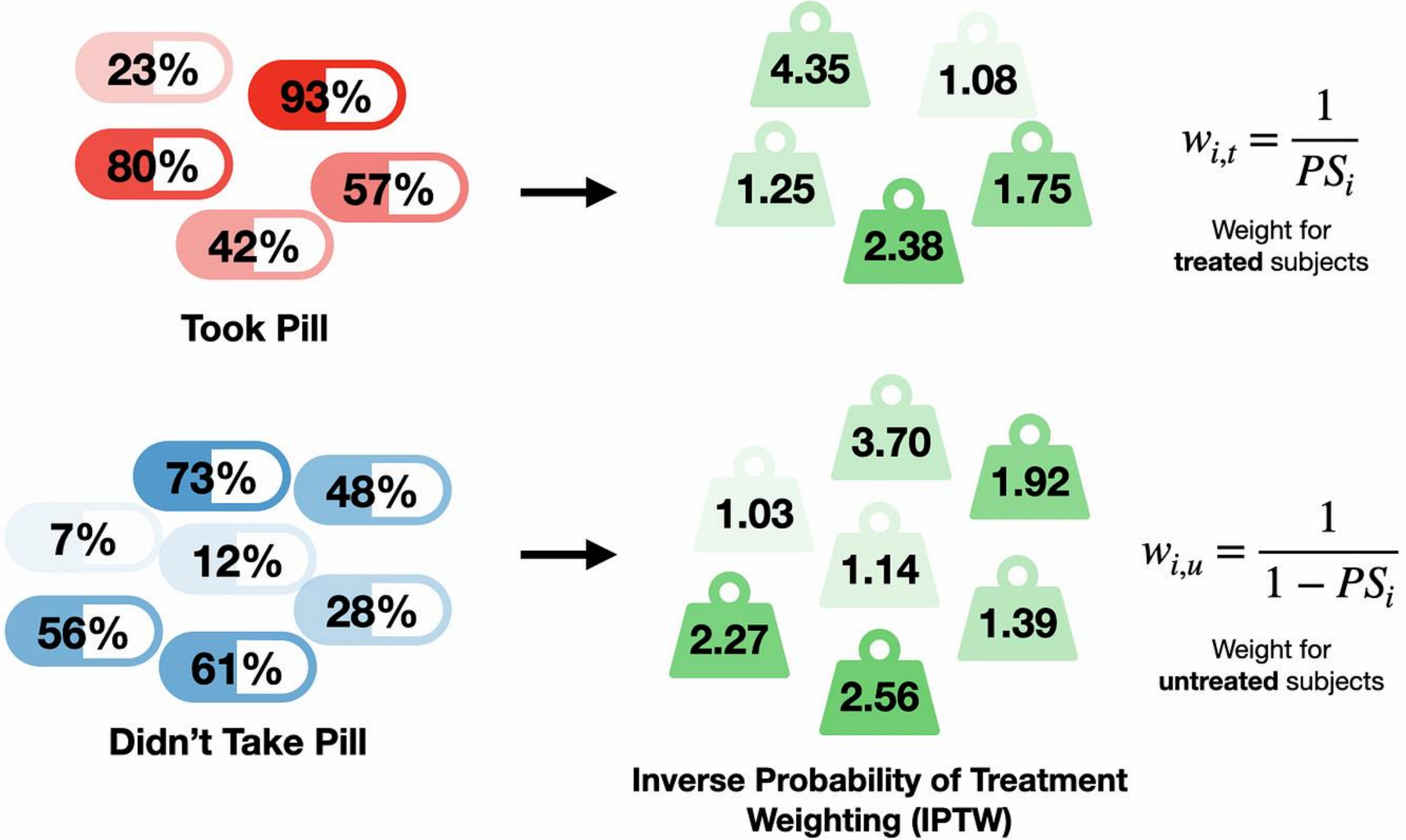
Estimate with machine learning!



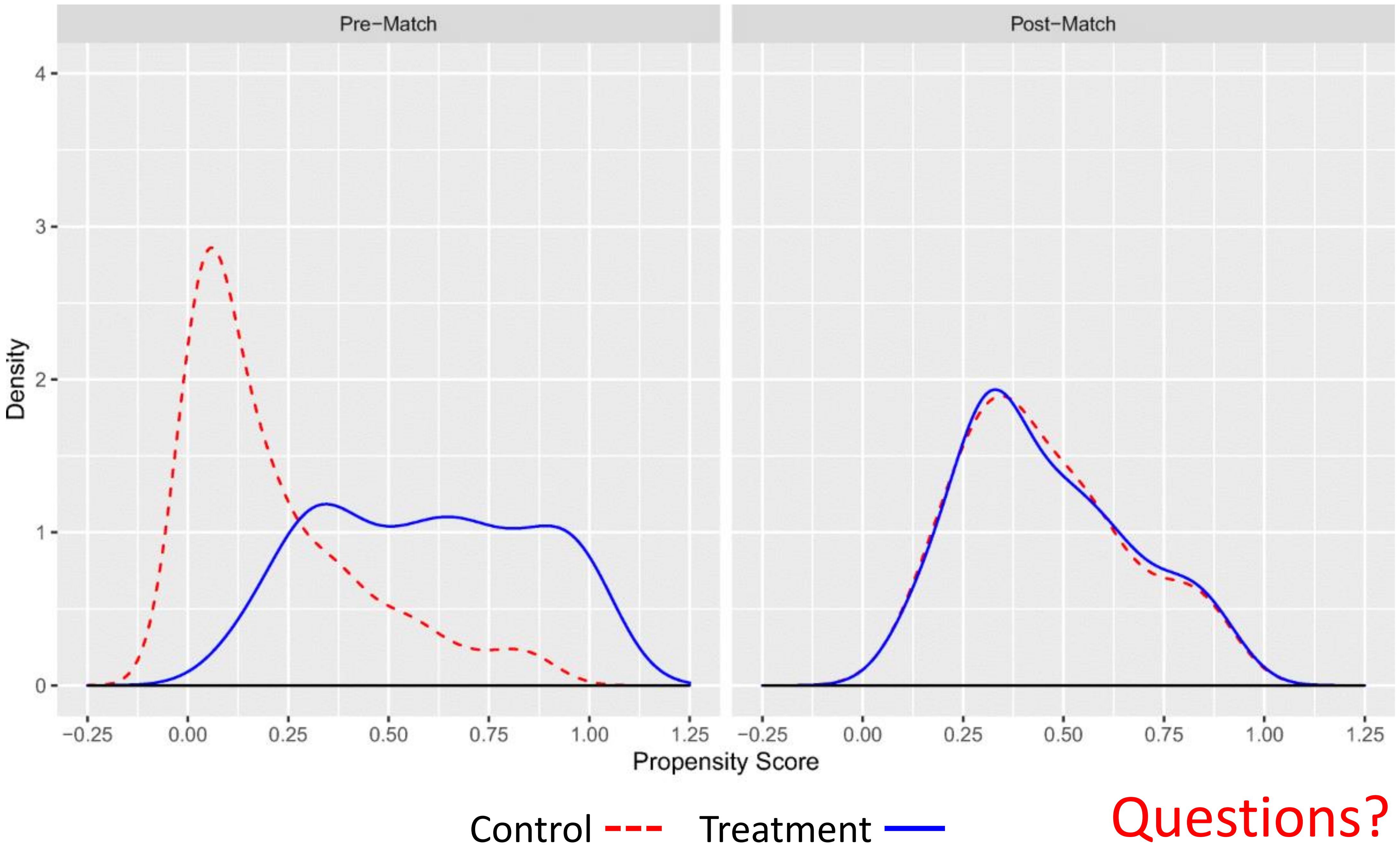

23%
93%
80%
57%
42%
Took Pill
73%
48%
7%
12%
28%
56%
61%
Didn't Take Pill
93%
80%
73%
61%
57%
56%
48%
42%
28%
23%
12%
7%
Group 1
Group 2
Group 3
Group 4
Stratification


23%
93%
80%
57%
42%
Took Pill
73%
48%
7%
12%
28%
56%
61%
Didn't Take Pill
93%
73%
80%
61%
57%
56%
42%
41%
23%
28%
Matching
(without replacement)


23%
93%
80%
57%
42%
Took Pill
73%
48%
7%
12%
28%
56%
61%
Didn't Take Pill
93%
73%
80%
73%
Same observation.
57%
56%
42%
41%
23%
28%
Matching
~~(without replacement)~~
(with replacement)

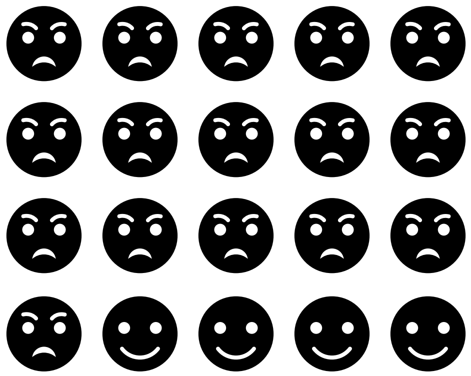


23%
93%
80%
57%
42%
Took Pill
4.35
1.08
1.25
1.75
2.38
$w_{i,t} = \frac{1}{PS_i}$
Weight for
treated subjects
73%
48%
7%
12%
28%
56%
61%
Didn't Take Pill
3.70
1.92
1.03
1.14
2.27
1.39
2.56
$w_{i,u} = \frac{1}{1 - PS_i}$
Weight for
untreated subjects
Inverse Probability of Treatment
Weighting (IPTW)


Pre−Match
Post−Match
Density
Propensity Score
4
3
2
1
0
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
Control
Treatment

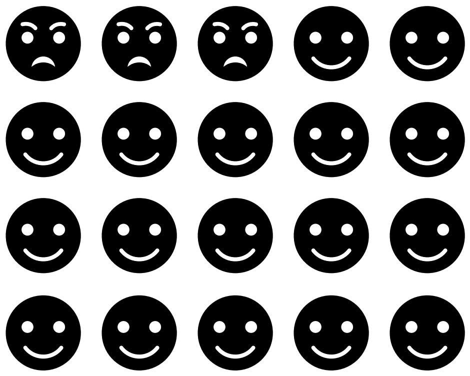
Questions?

Incentive



Churn 30%

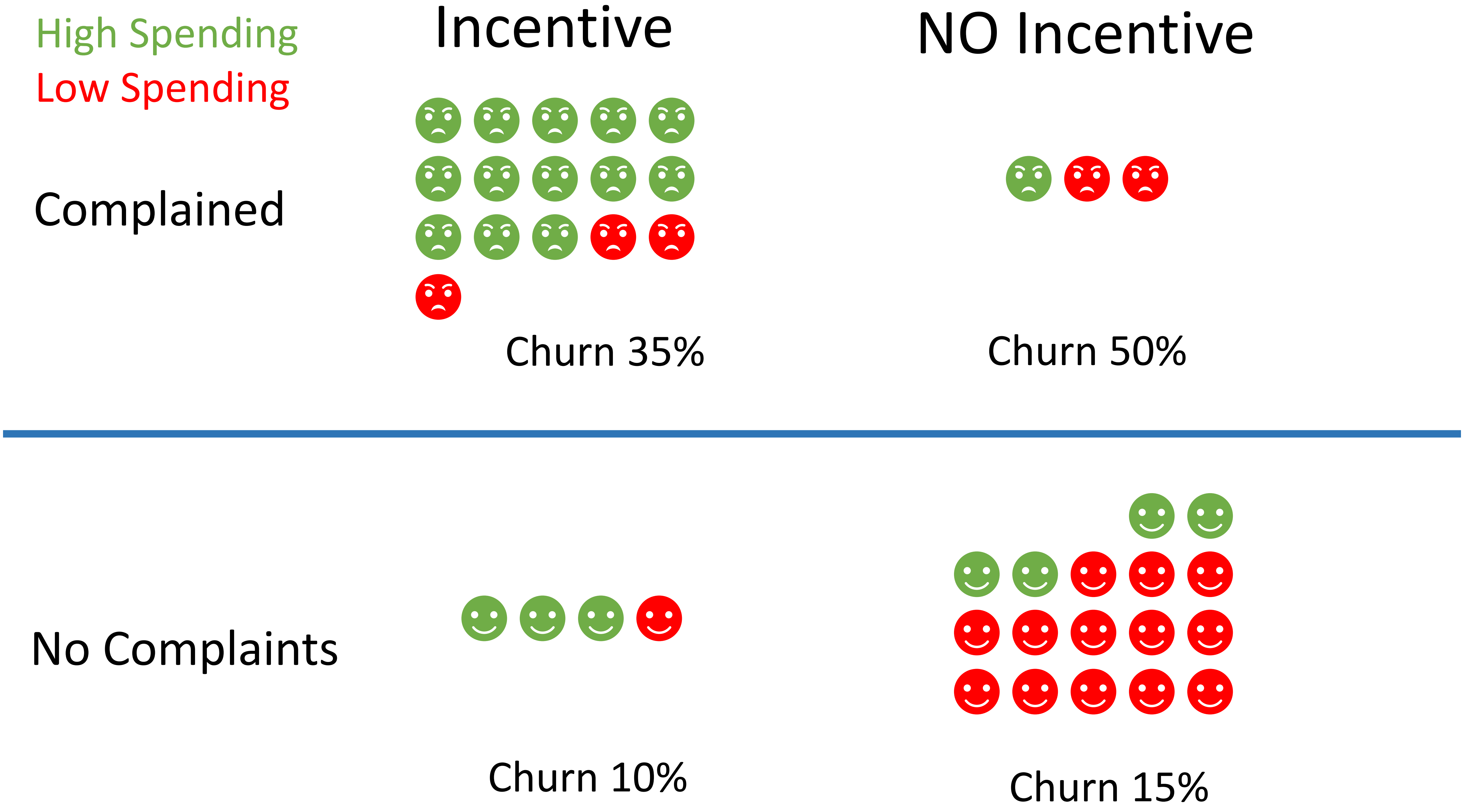
No Incentive



Churn 20%

Concerns?

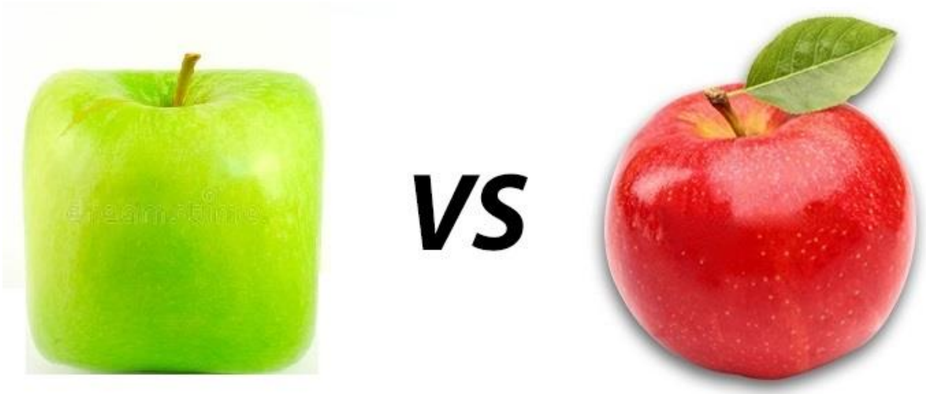
	Incentive	NO Incentive
Complained	Churn 35%	Churn 50%
No Complaints	Churn 10%	Churn 15%


High Spending
Low Spending
Incentive
NO Incentive
Complained
Churn 35%
Churn 50%
No Complaints
Churn 10%
Churn 15%

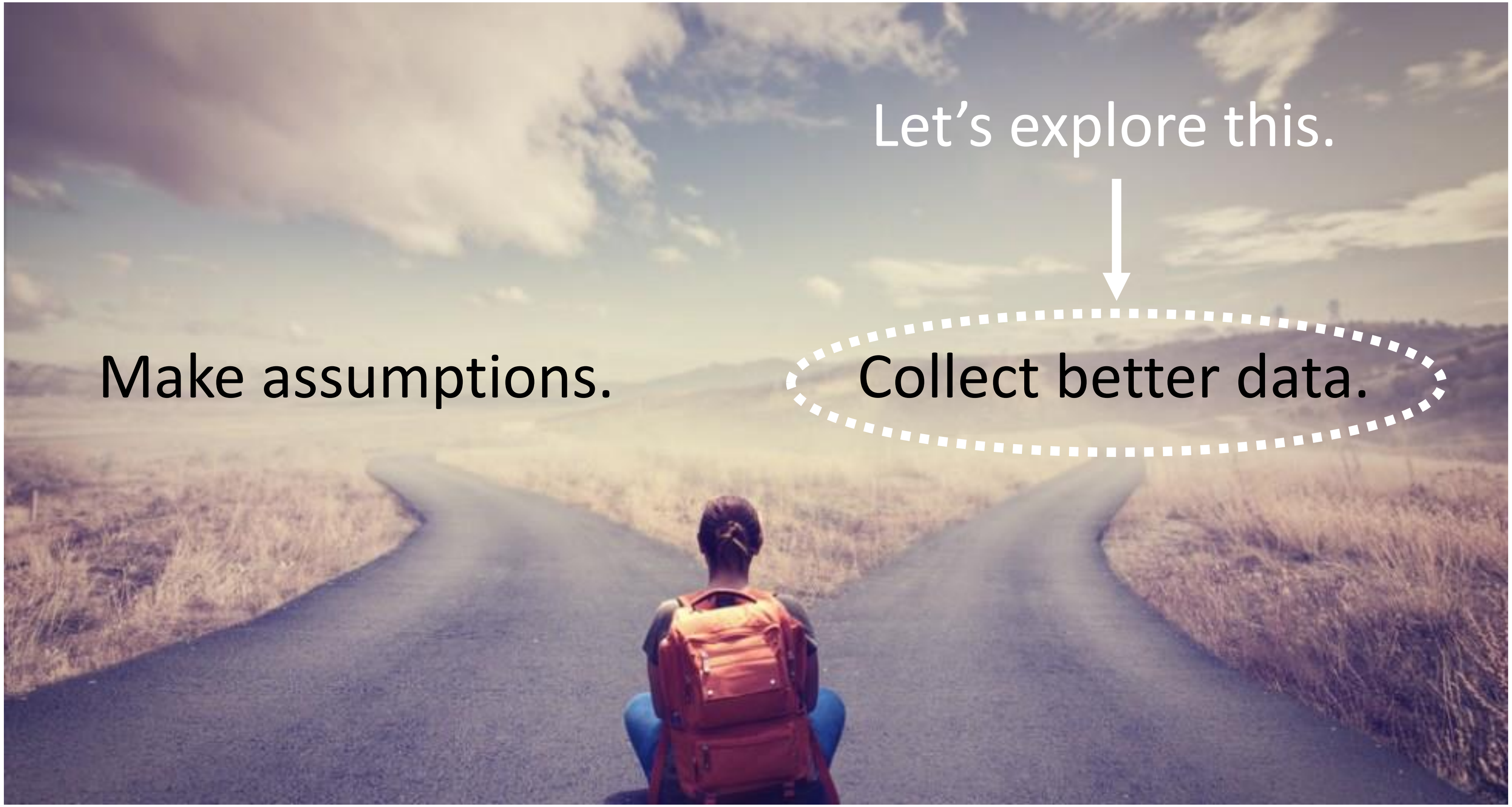
Key Takeaway

We can’t be certain that groups are comparable by just looking at the data.

What you may actually be doing:

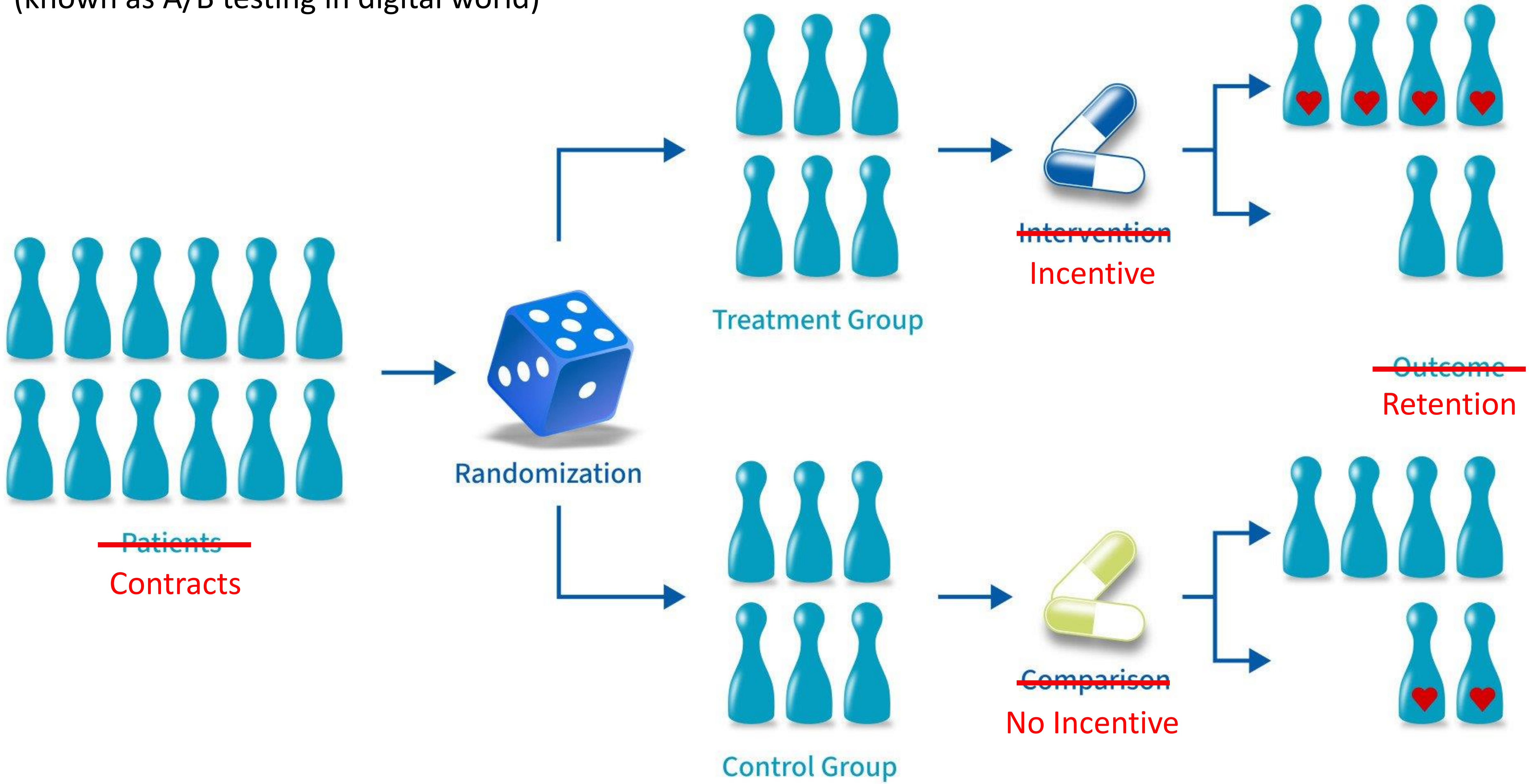


Our options are:



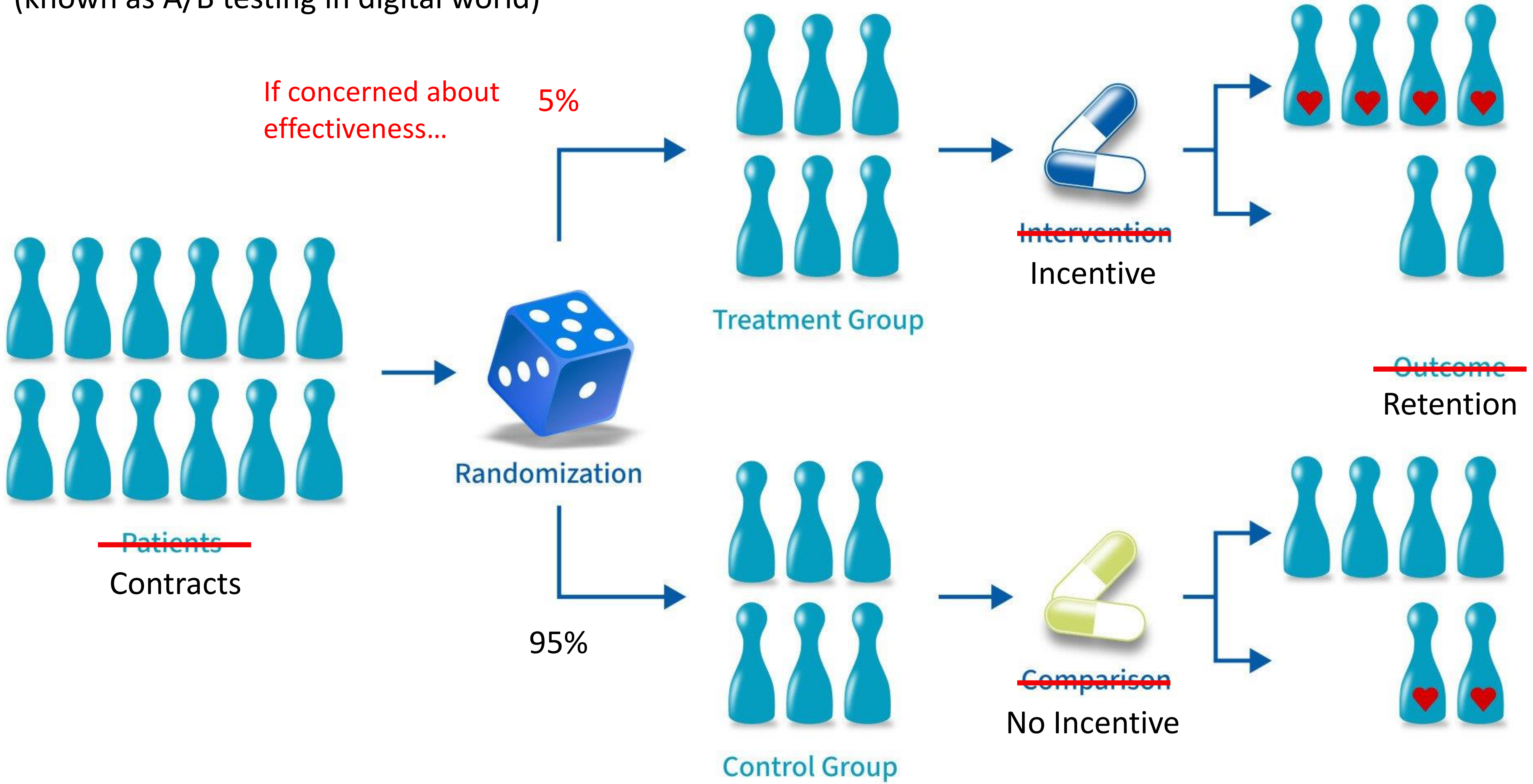
Randomized Controlled Trial

(known as A/B testing in digital world)



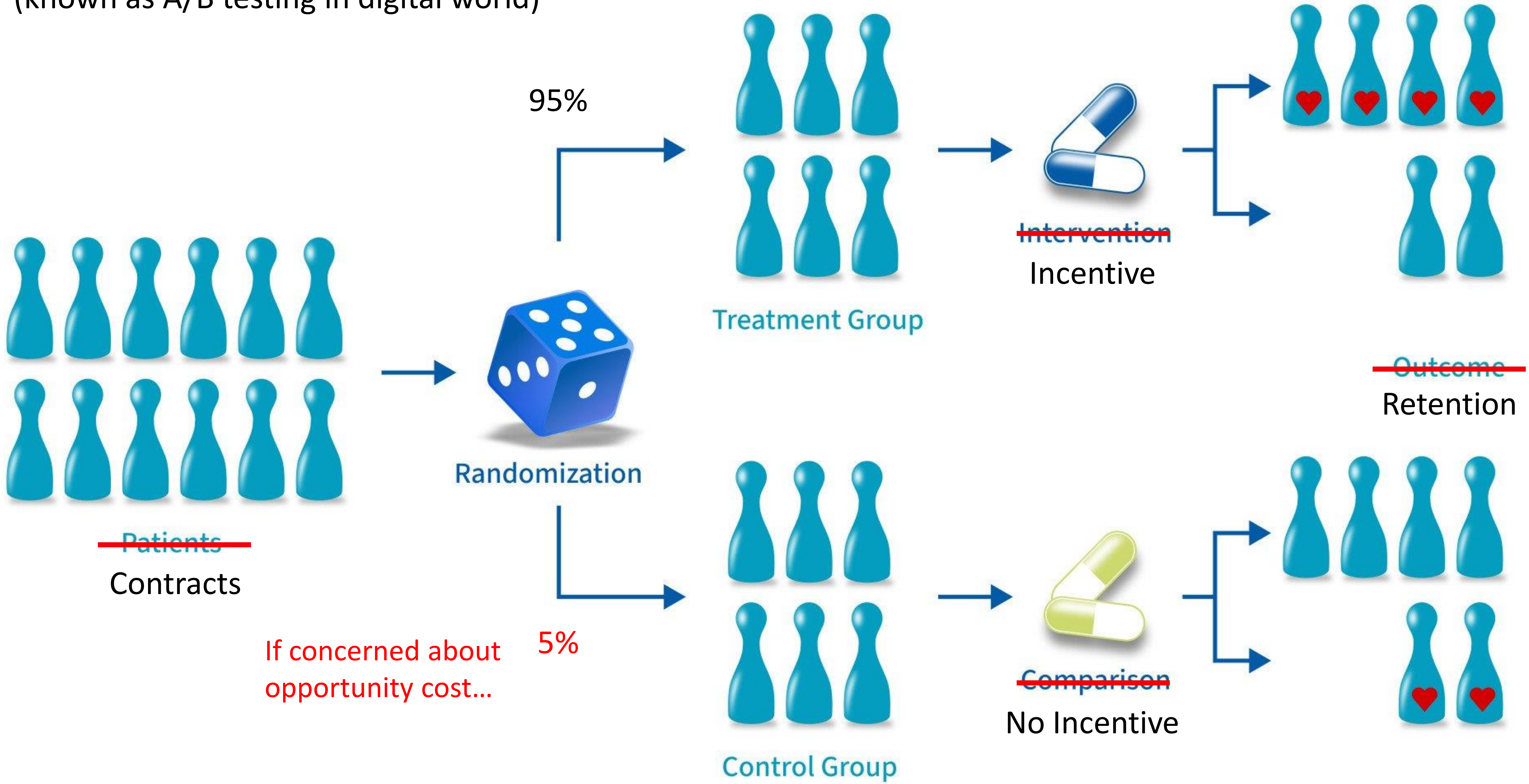
Randomized Controlled Trial

(known as A/B testing in digital world)



Randomized Controlled Trial

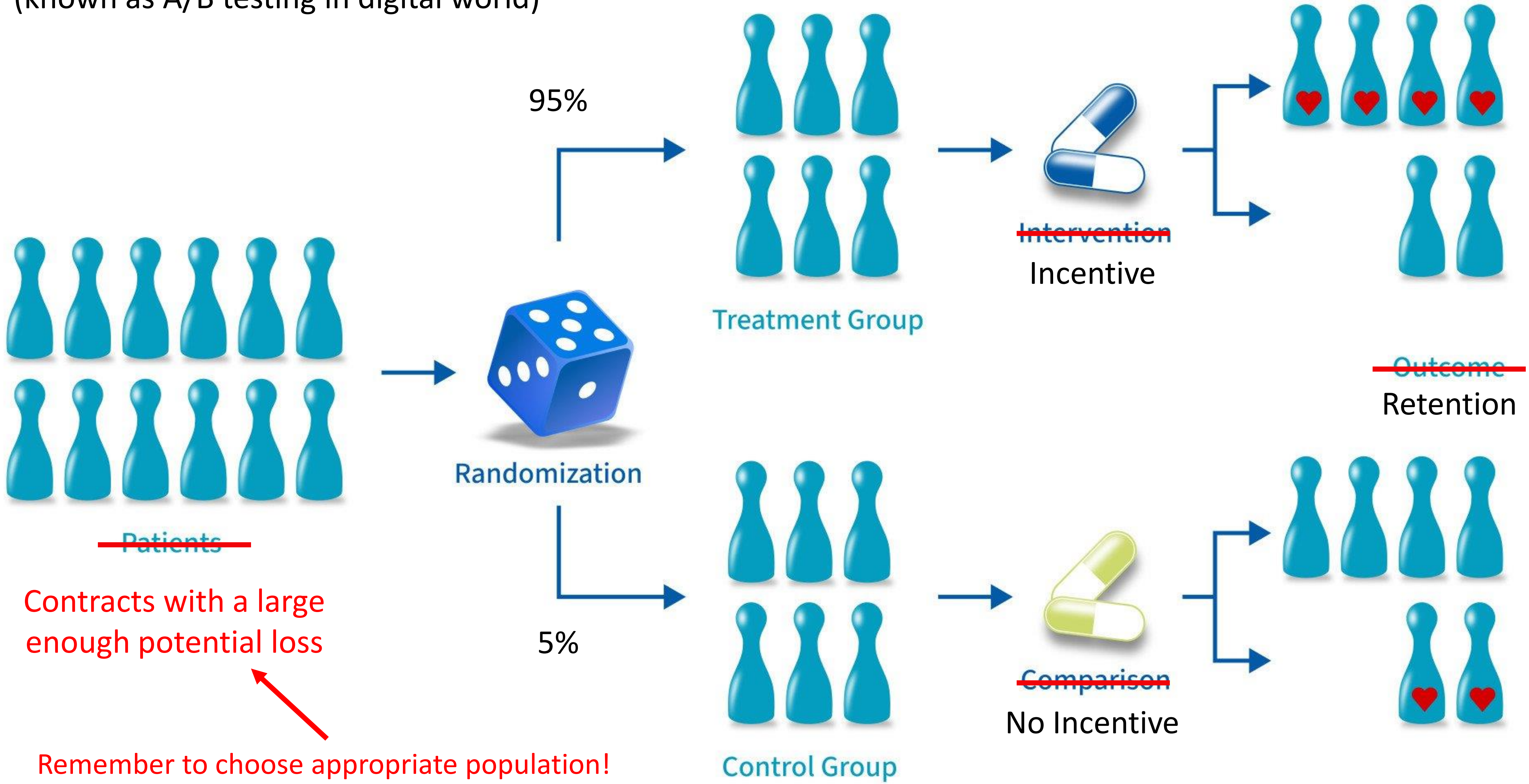
(known as A/B testing in digital world)



If concerned about opportunity cost…

Randomized Controlled Trial

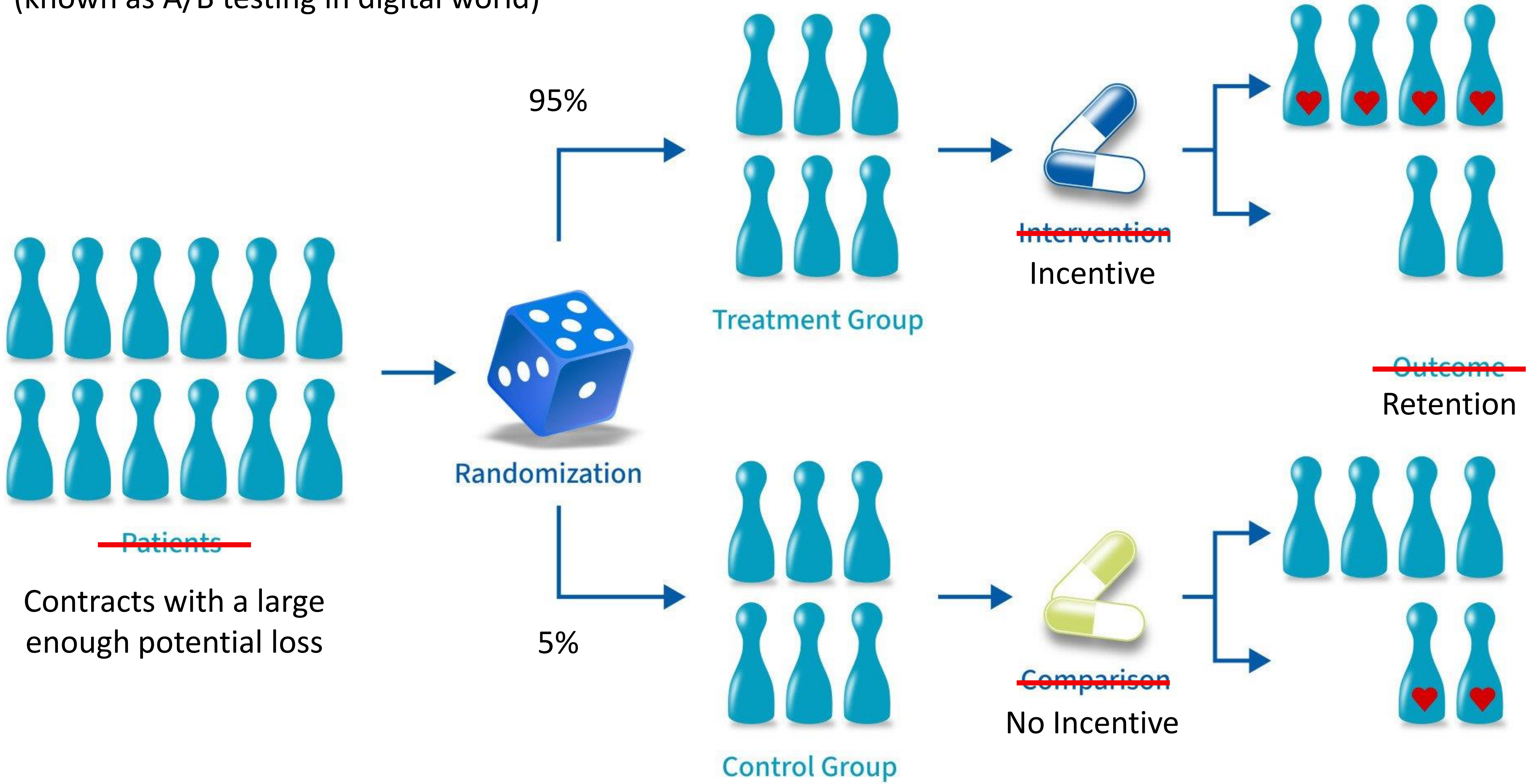
(known as A/B testing in digital world)

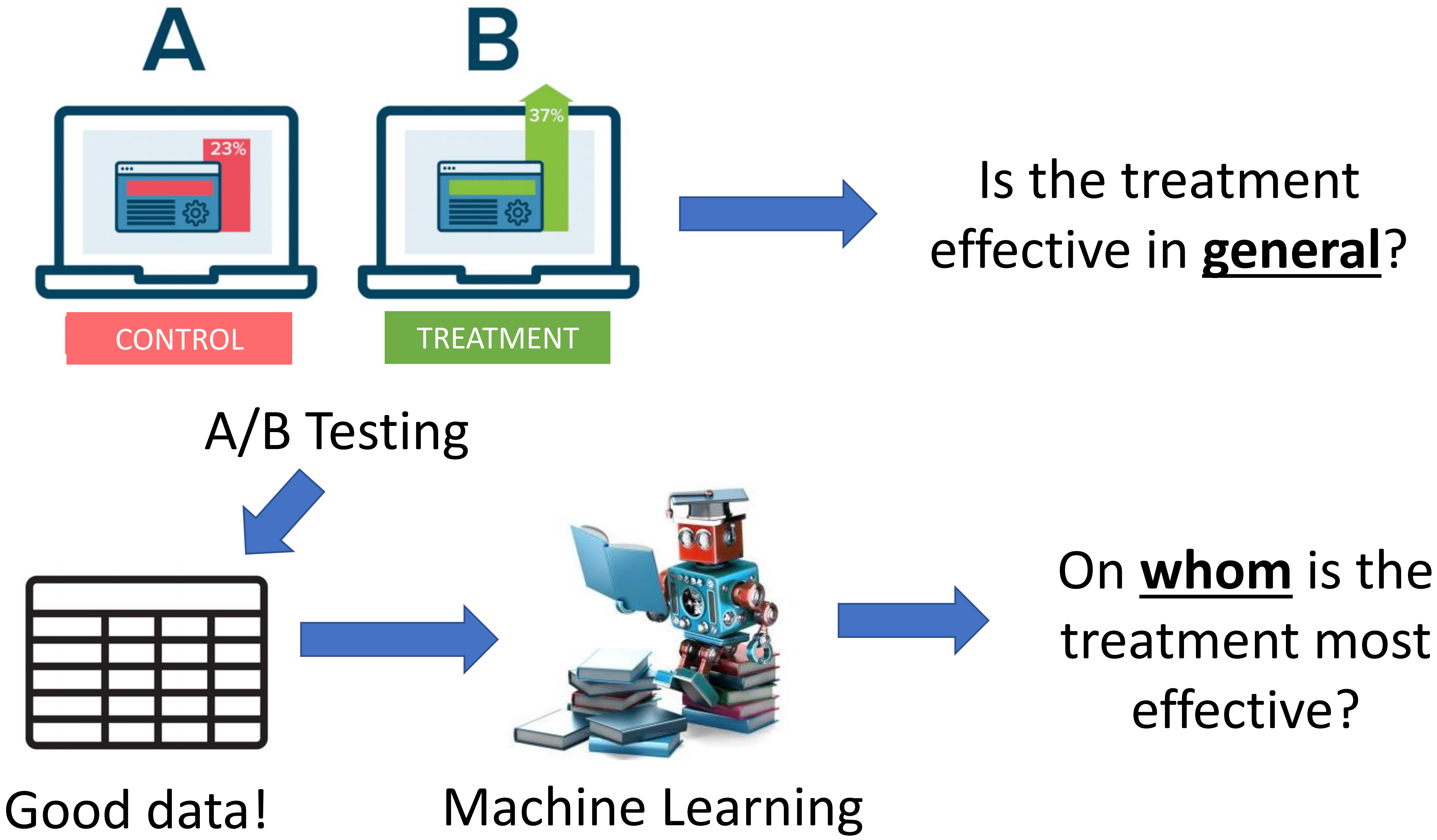


Randomized Controlled Trial

(known as A/B testing in digital world)

Questions?

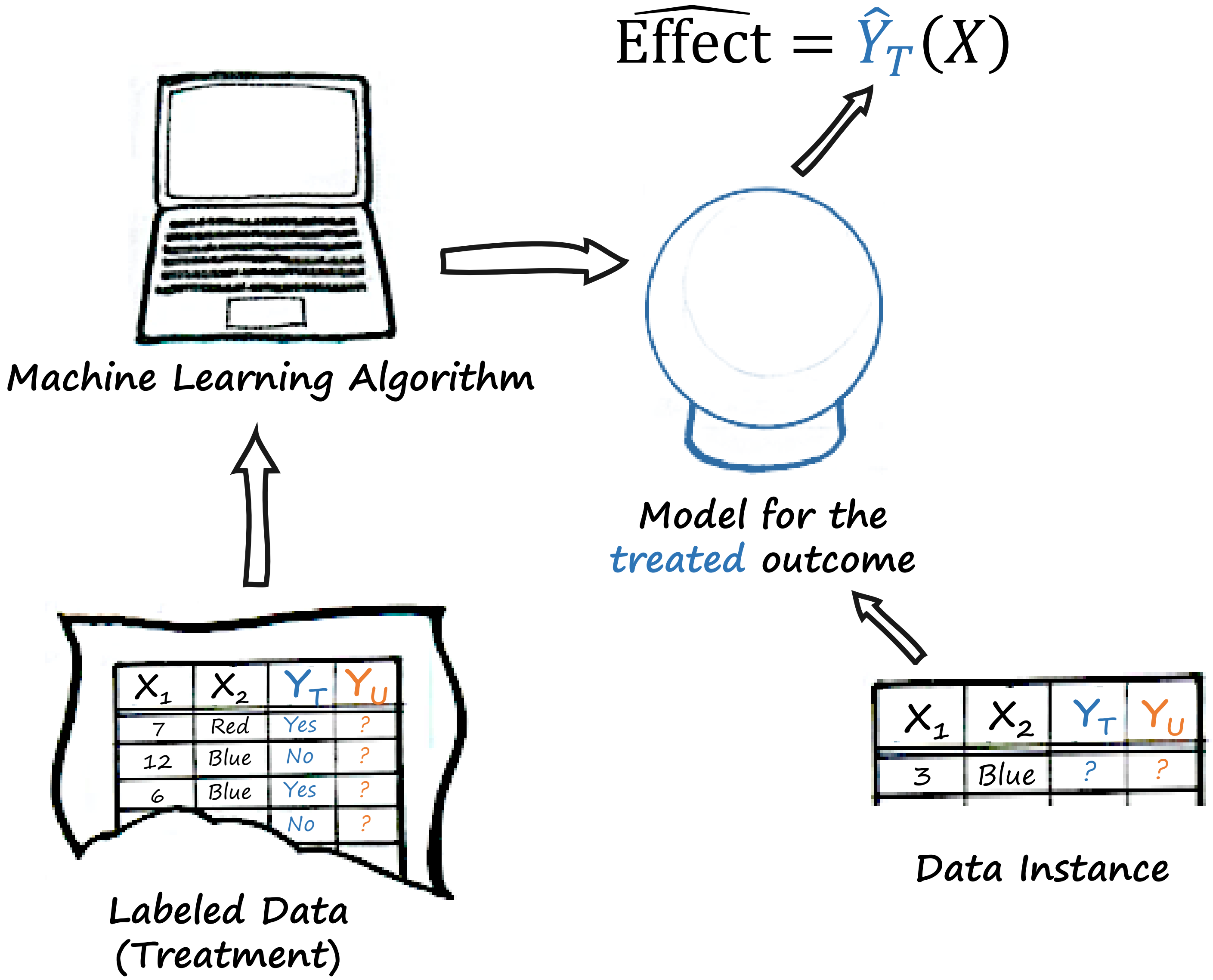



A
B
23%
37%
CONTROL
TREATMENT
A/B Testing
Is the treatment effective in general?
Good data!
Machine Learning
On whom is the treatment most effective?

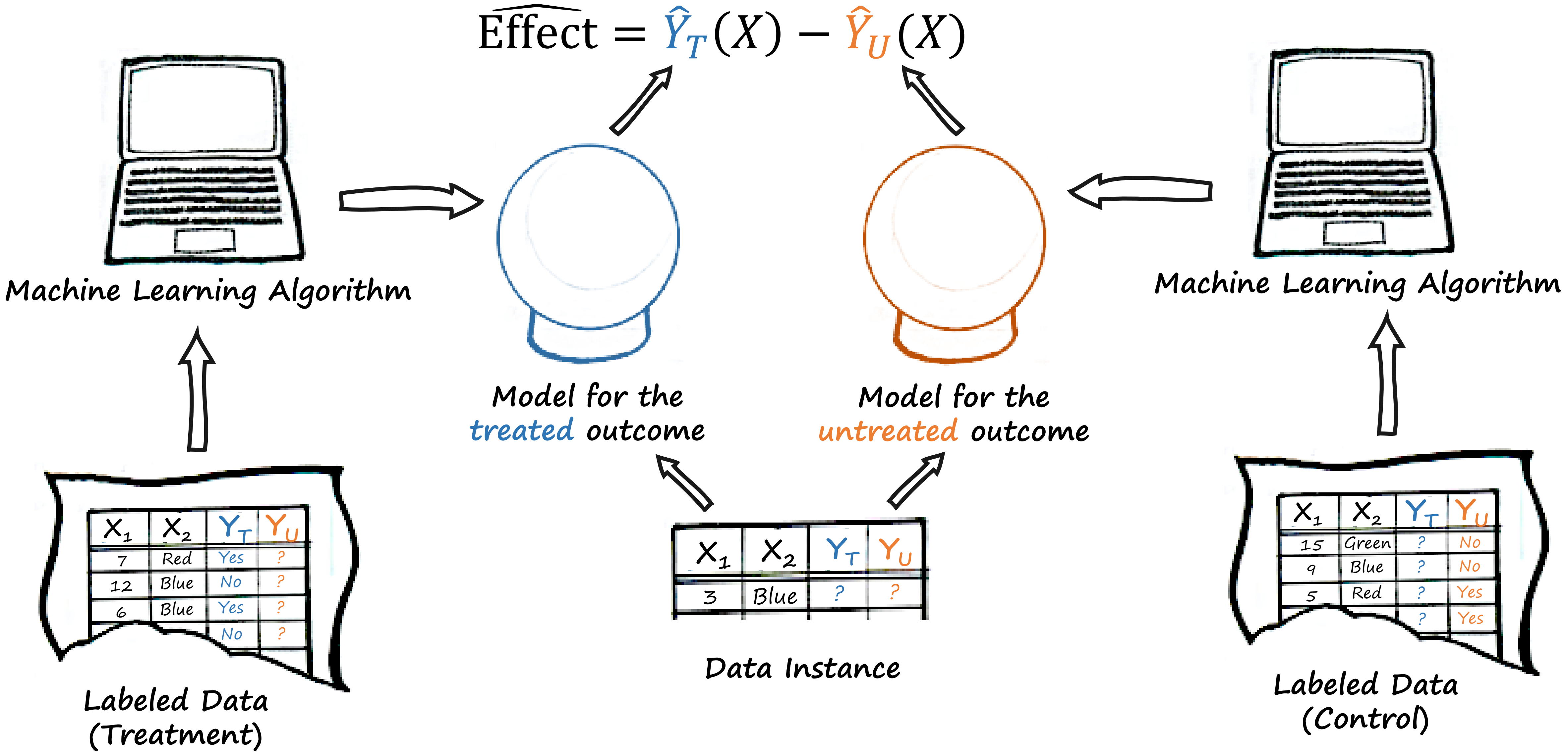
$$\text{Effect} = Y_T - Y_U$$

Incentive NO Incentive

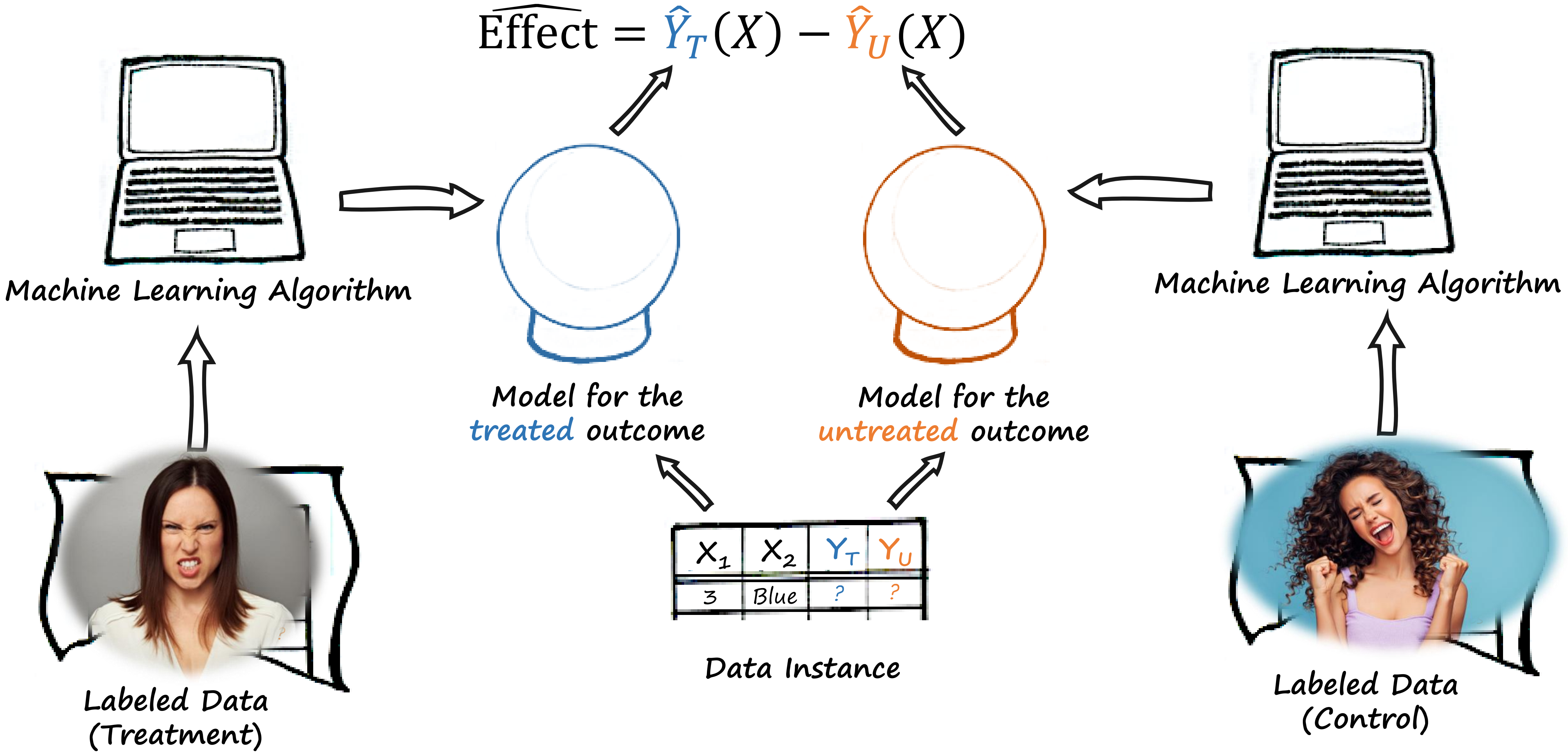
Y_T and Y_U are the potential outcomes if treated and untreated.


$\widehat{\text{Effect}} = \hat{Y}_T(X)$
Machine Learning Algorithm
Model for the treated outcome
Labeled Data (Treatment)
Data Instance
X_1 X_2 Y_T Y_U
7 Red Yes ?
12 Blue No ?
6 Blue Yes ?
No ?
3 Blue ? ?

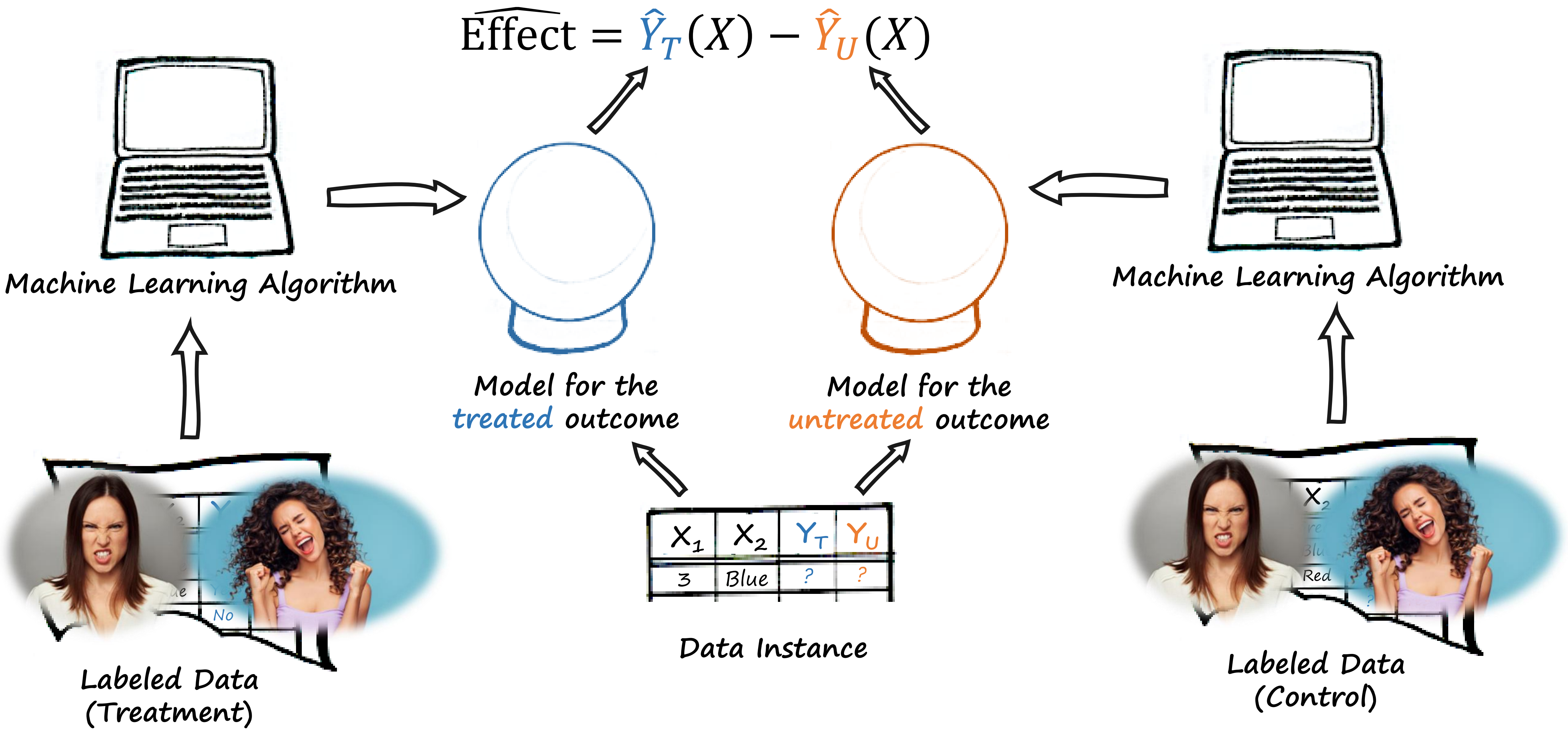
Questions?


$\widehat{\text{Effect}} = \hat{Y}_T(X) - \hat{Y}_U(X)$
Machine Learning Algorithm
Model for the treated outcome
Model for the untreated outcome
Machine Learning Algorithm
X1 X2 YT YU
7 Red Yes ?
12 Blue No ?
6 Blue Yes ?
No ?
Labeled Data (Treatment)
X1 X2 YT YU
3 Blue ? ?
Data Instance
X1 X2 YT YU
15 Green ? No
9 Blue ? No
5 Red ? Yes
? Yes
Labeled Data (Control)

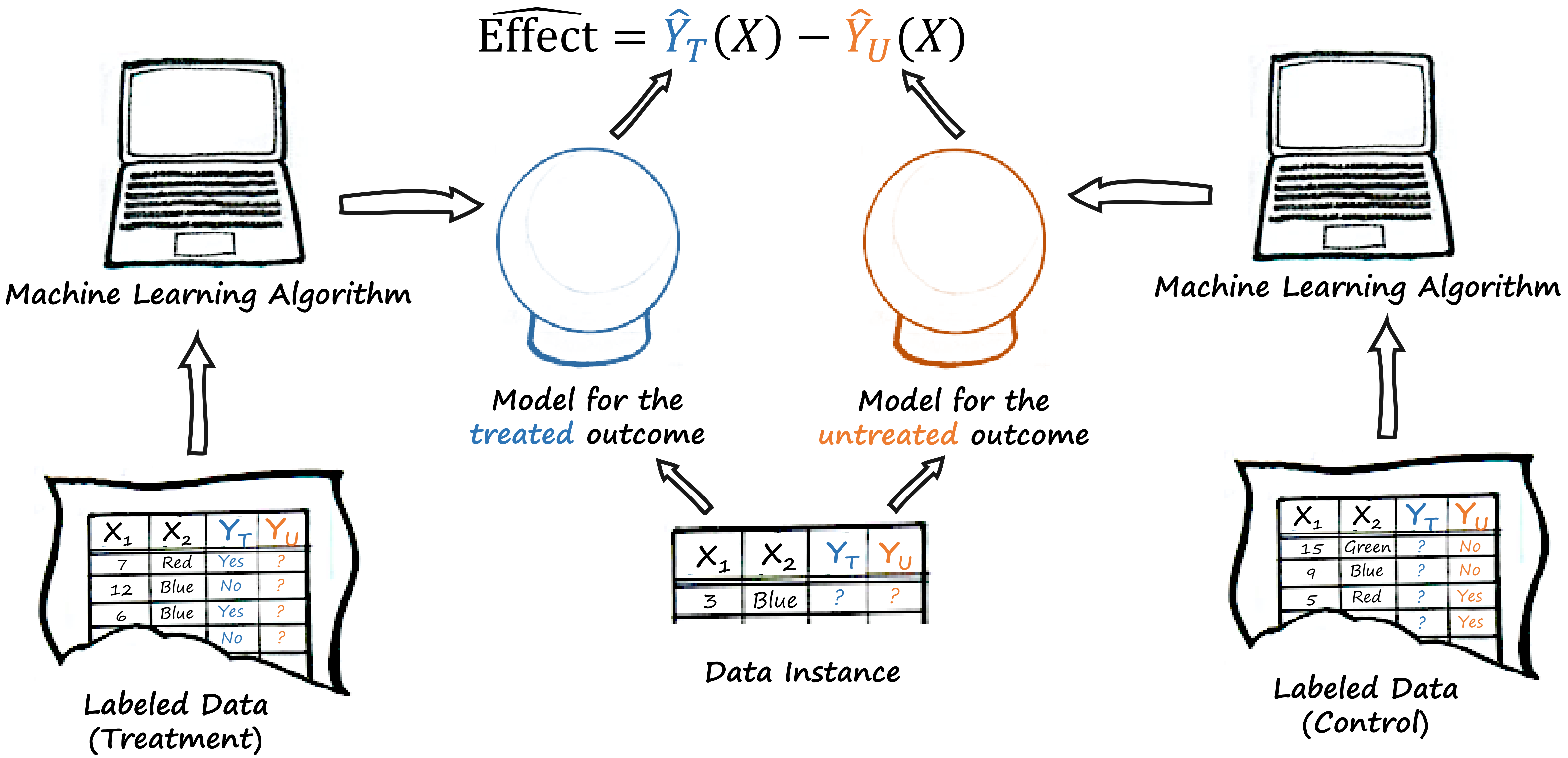
Beware of confounding!



That’s why we use data from an A/B test



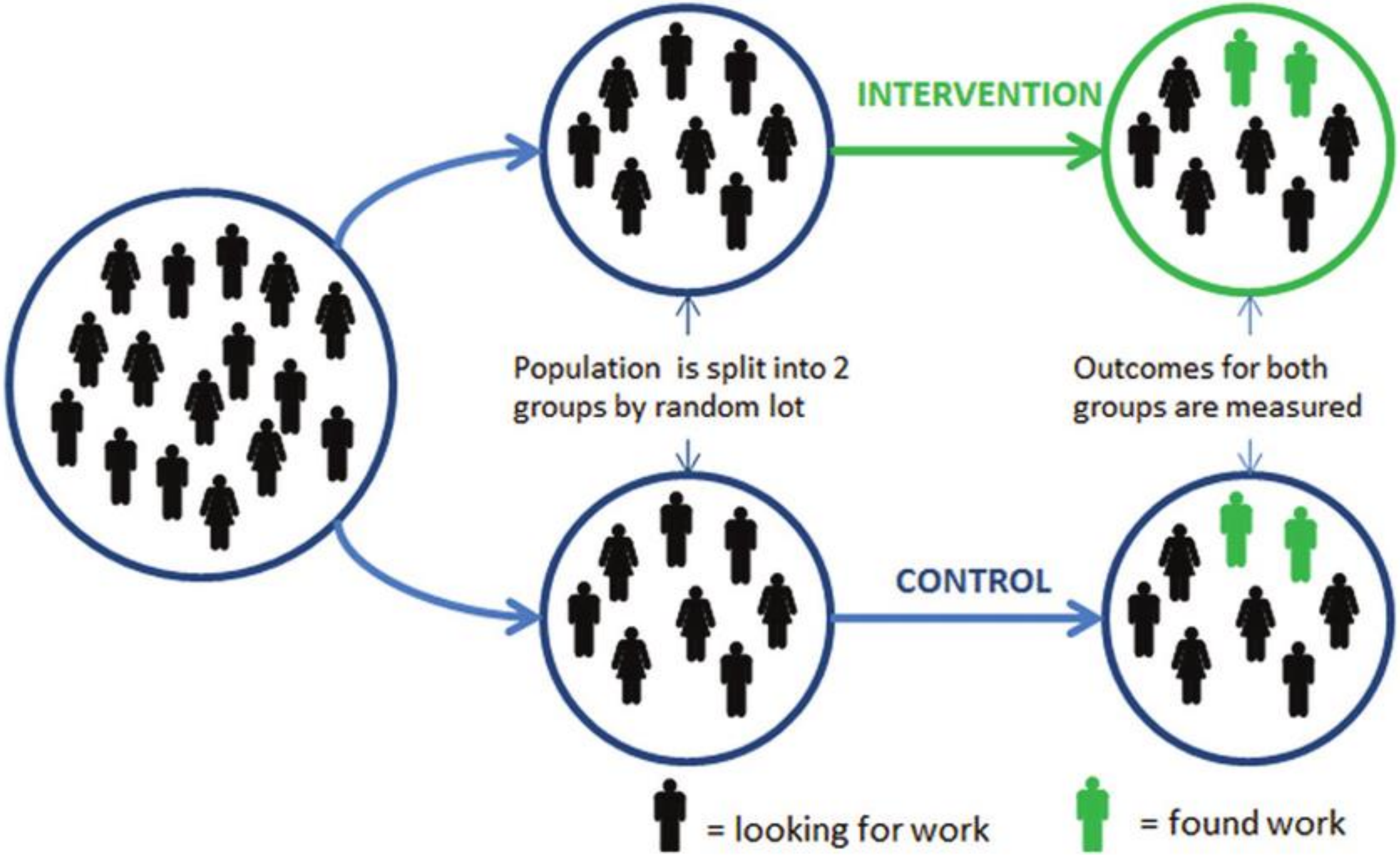
Questions?



$\widehat{\text{Effect}} = \hat{Y}_T(X) - \hat{Y}_U(X)$
Machine Learning Algorithm
Model for the treated outcome
Model for the untreated outcome
Machine Learning Algorithm
X_1 X_2 Y_T Y_U
7 Red Yes ?
12 Blue No ?
6 Blue Yes ?
No ?
Labeled Data (Treatment)
X_1 X_2 Y_T Y_U
3 Blue ? ?
Data Instance
X_1 X_2 Y_T Y_U
15 Green ? No
9 Blue ? No
5 Red ? Yes
? Yes
Labeled Data (Control)

Key Takeaway

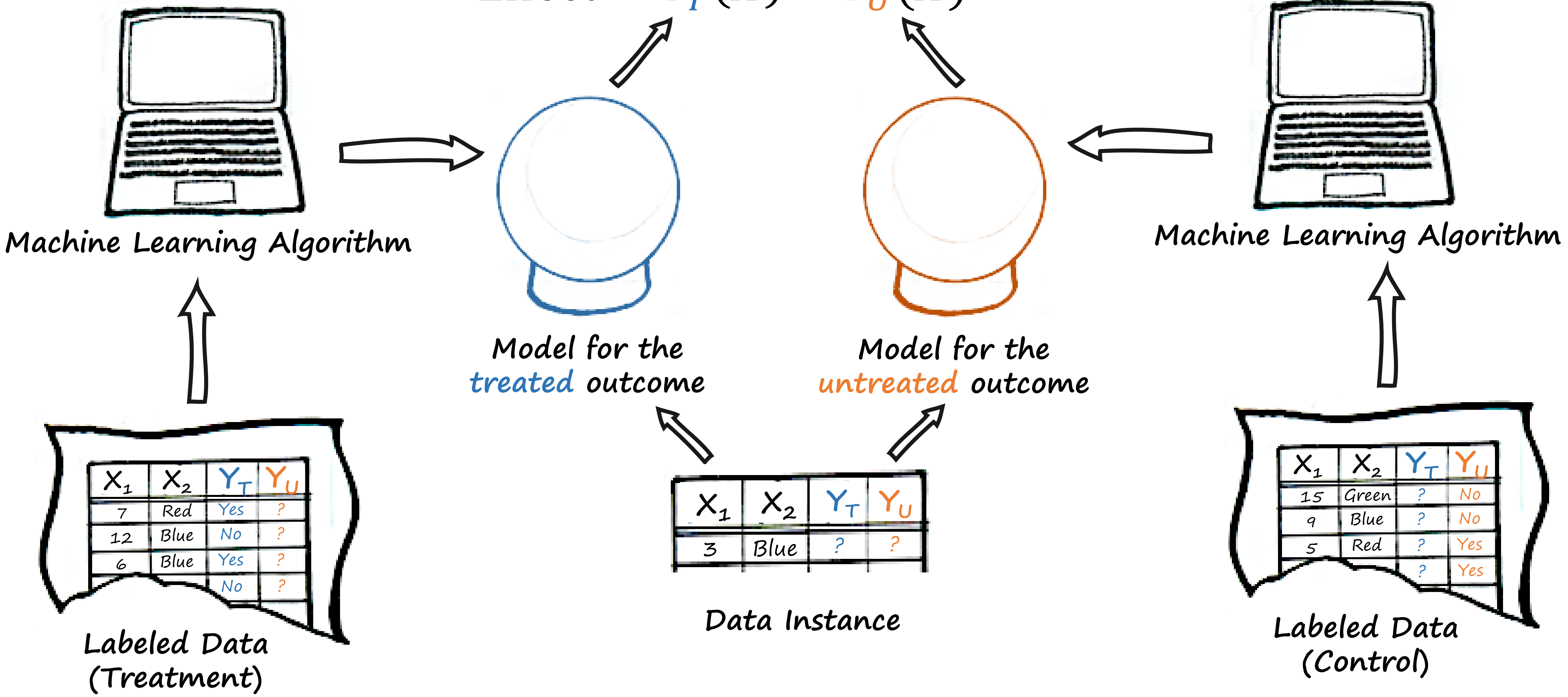
We can use A/B testing to collect data on comparable groups.



Key Takeaway

We can use Machine Learning + A/B Test data to predict how effects vary.

$$\widehat{\text{Effect}} = \hat{Y}_T(X) - \hat{Y}_U(X)$$



Evaluating A Decision Rule with A/B Test Data

Effect of Decision Rule

$$N \times \mathbb{P}[D = 1] \times (\mathbb{E}[Value|D = 1, T = 1] - \mathbb{E}[Value|D = 1, T = 0])$$

- T: Treatment condition in the A/B Test Data; 1 = Treat, 0 = Control
- D: Decision made with predictive models; 1 = Treat, 0 = Control
- N: Number of decisions to be made.
- $Value$: Based on benefit matrix.

Effect of Decision Rule

$$N \times \mathbb{P}[D = 1] \times (\mathbb{E}[Value|D = 1, T = 1] - \mathbb{E}[Value|D = 1, T = 0])$$

Average value when not treated (estimate from cases where treatment is decided but not applied).

Average value when treated (estimate from cases where treatment is decided and applied).

Fraction of treat decisions (average of D).

Key Takeaway

We can also use A/B test data to evaluate causal decisions.

$$N \times \mathbb{P}[D = 1] \times \underbrace{(\mathbb{E}[Value|D = 1, T = 1] - \mathbb{E}[Value|D = 1, T = 0])}_{\text{Average effect for those treated by the decision rule.}}$$

Fraction treated by the decision rule.

The Plan

~~Fundamentals~~

Week 1: ~~Problem formulation.~~

Week 2: ~~Modeling (Part 1).~~

Week 3: ~~Modeling (Part 2).~~

Week 4: ~~Model Evaluation.~~

(What I think are)

The Big Three

Week 5: ~~Causal analytics.~~

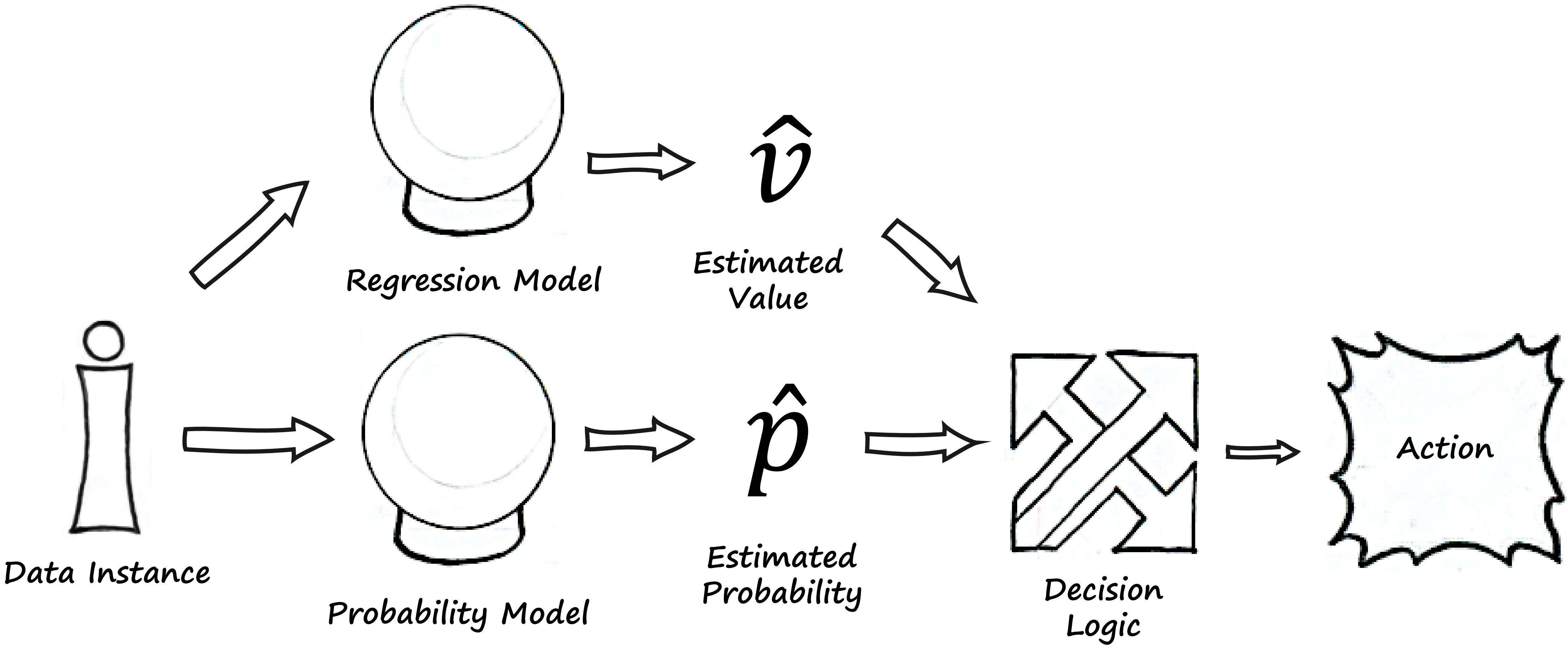
Week 6: Big (unstructured) data.

Week 7: Generative AI (e.g., ChatGPT).

Week 8: Final Exam.

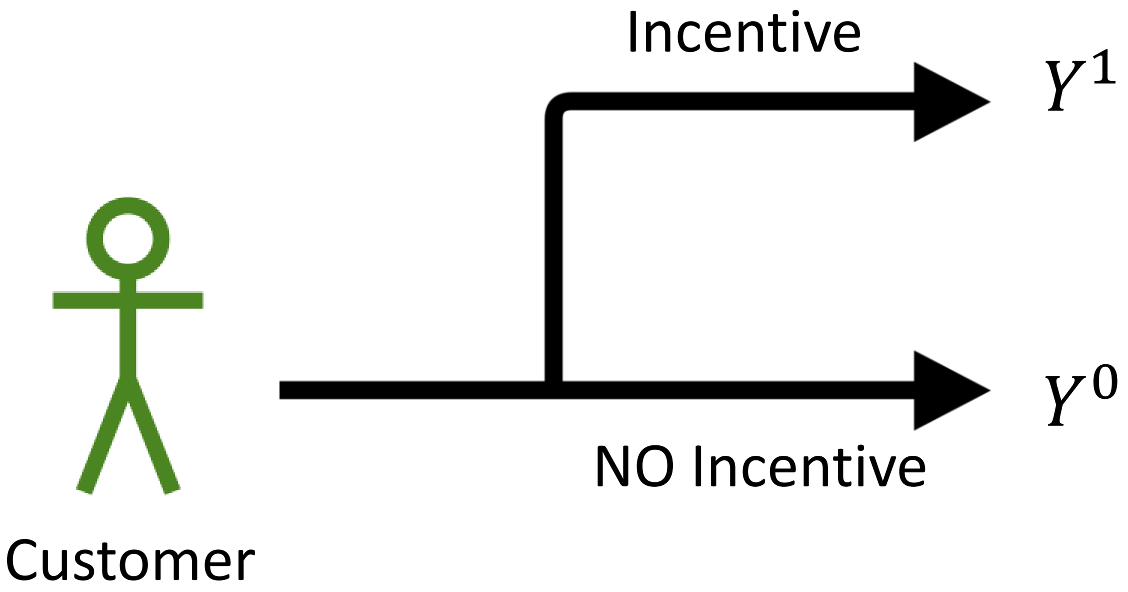
Key Takeaway

Our decision logic can incorporate multiple quantities
(possibly from multiple models).



Key Takeaway

A causal effect is the difference between two **potential** outcomes.



$$\text{Effect} = Y^1 - Y^0$$

Key Takeaway

We can’t observe both potential outcomes. So:

We estimate effects by comparing people exposed to different conditions.

Give Incentive

Do NOT Give Incentive

Key Takeaway

Beware of treatments that could affect multiple people.

Key Takeaway

To infer causality, we must make apples-to-apples comparisons.

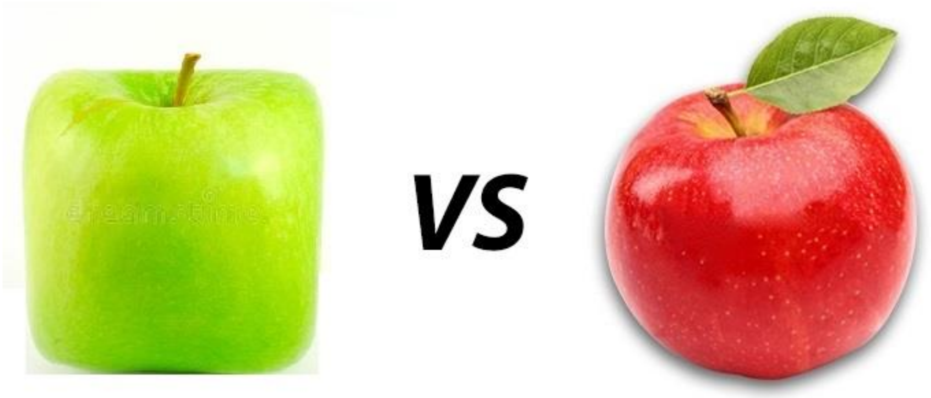
Churn rate with incentive − Churn rate without incentive ≠ Incentive Effect

Churn rate with incentive − Churn rate without incentive = Incentive Effect

Key Takeaway

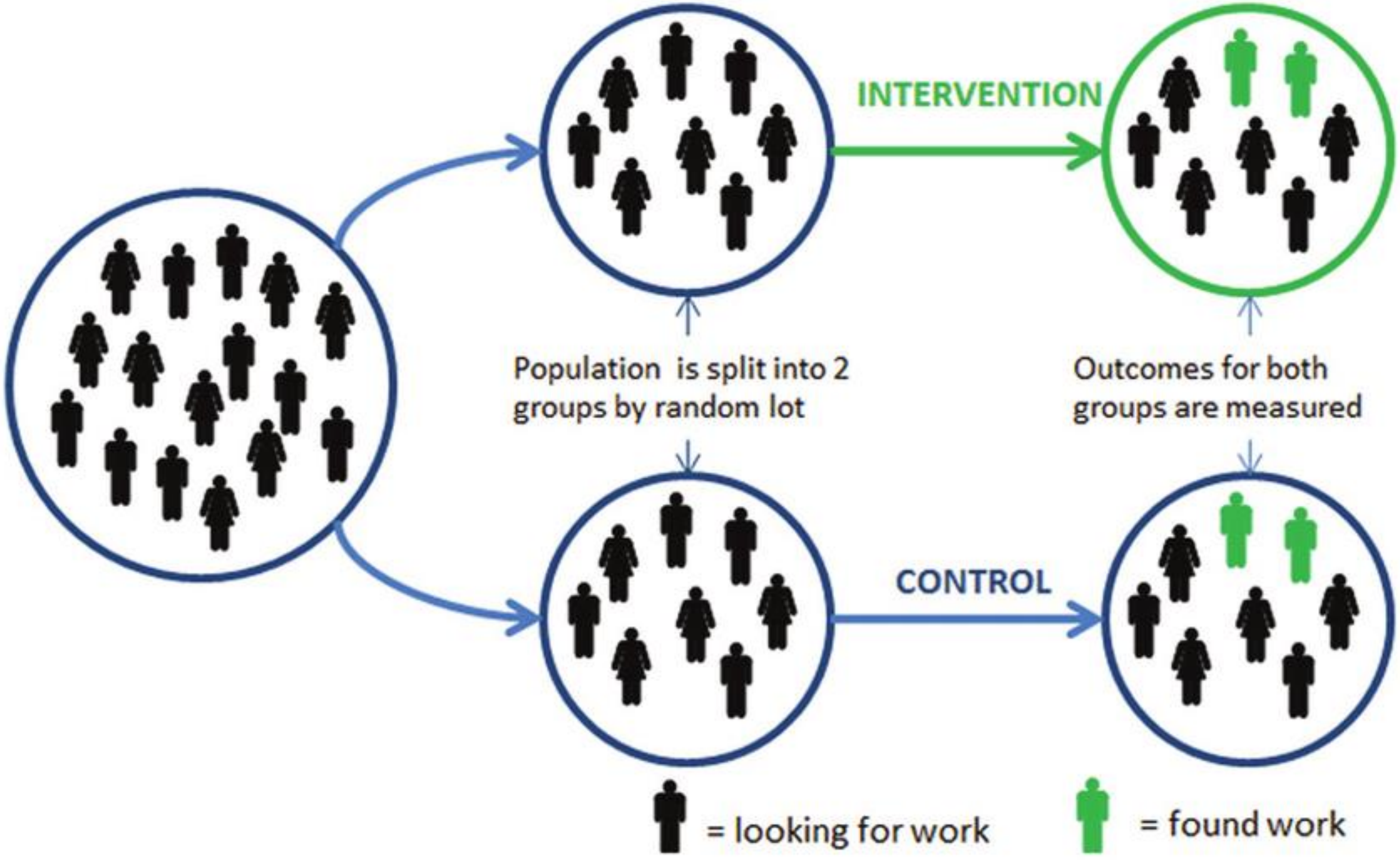
We can’t be certain that groups are comparable by just looking at the data.

What you may actually be doing:



Key Takeaway

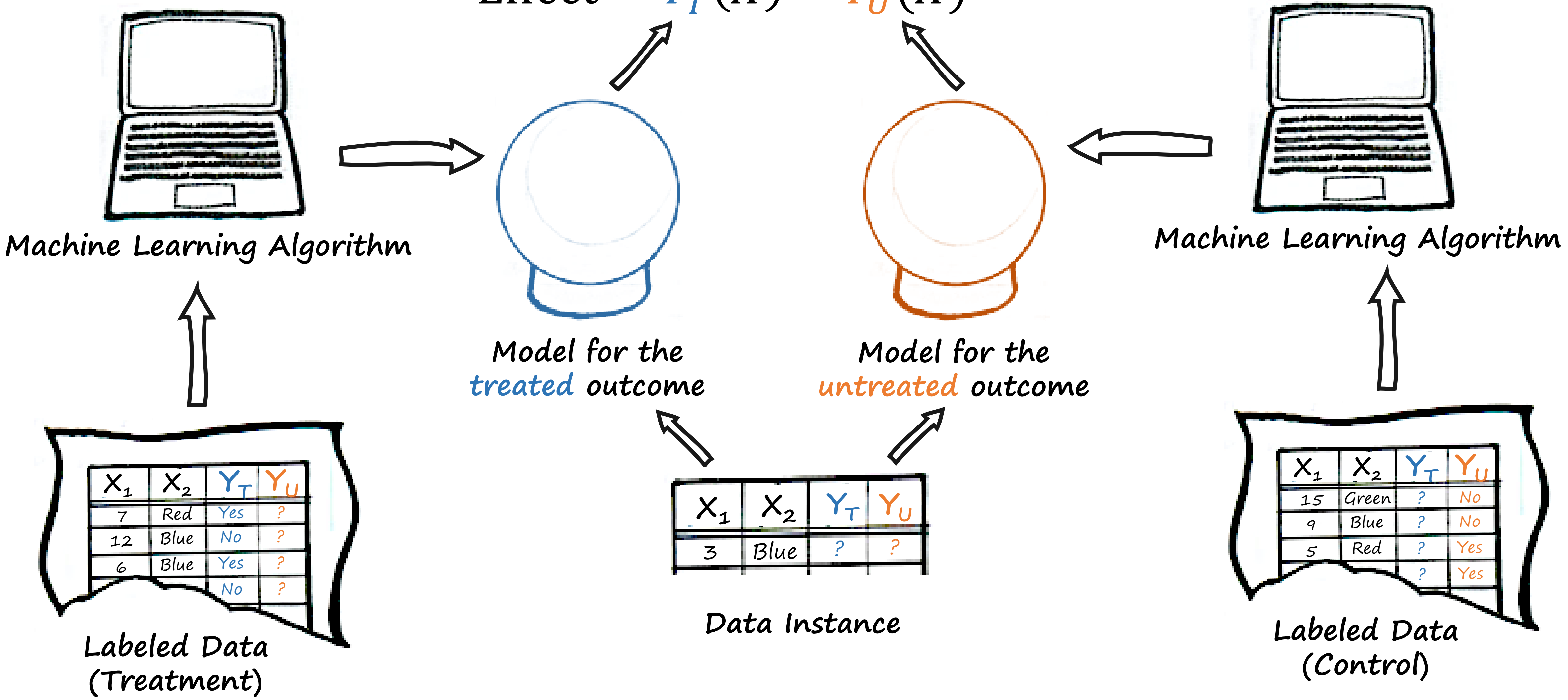
We can use A/B testing to collect data on comparable groups.



Key Takeaway

We can use Machine Learning + A/B Test data to predict how effects vary.

$$\widehat{\text{Effect}} = \hat{Y}_T(X) - \hat{Y}_U(X)$$



Key Takeaway

We can also use A/B test data to evaluate causal decisions.

$$N \times \mathbb{P}[D = 1] \times \underbrace{(\mathbb{E}[Value|D = 1, T = 1] - \mathbb{E}[Value|D = 1, T = 0])}_{\text{Average effect for those treated by the decision rule.}}$$

Fraction treated by the decision rule.

Average effect for those treated by the decision rule.