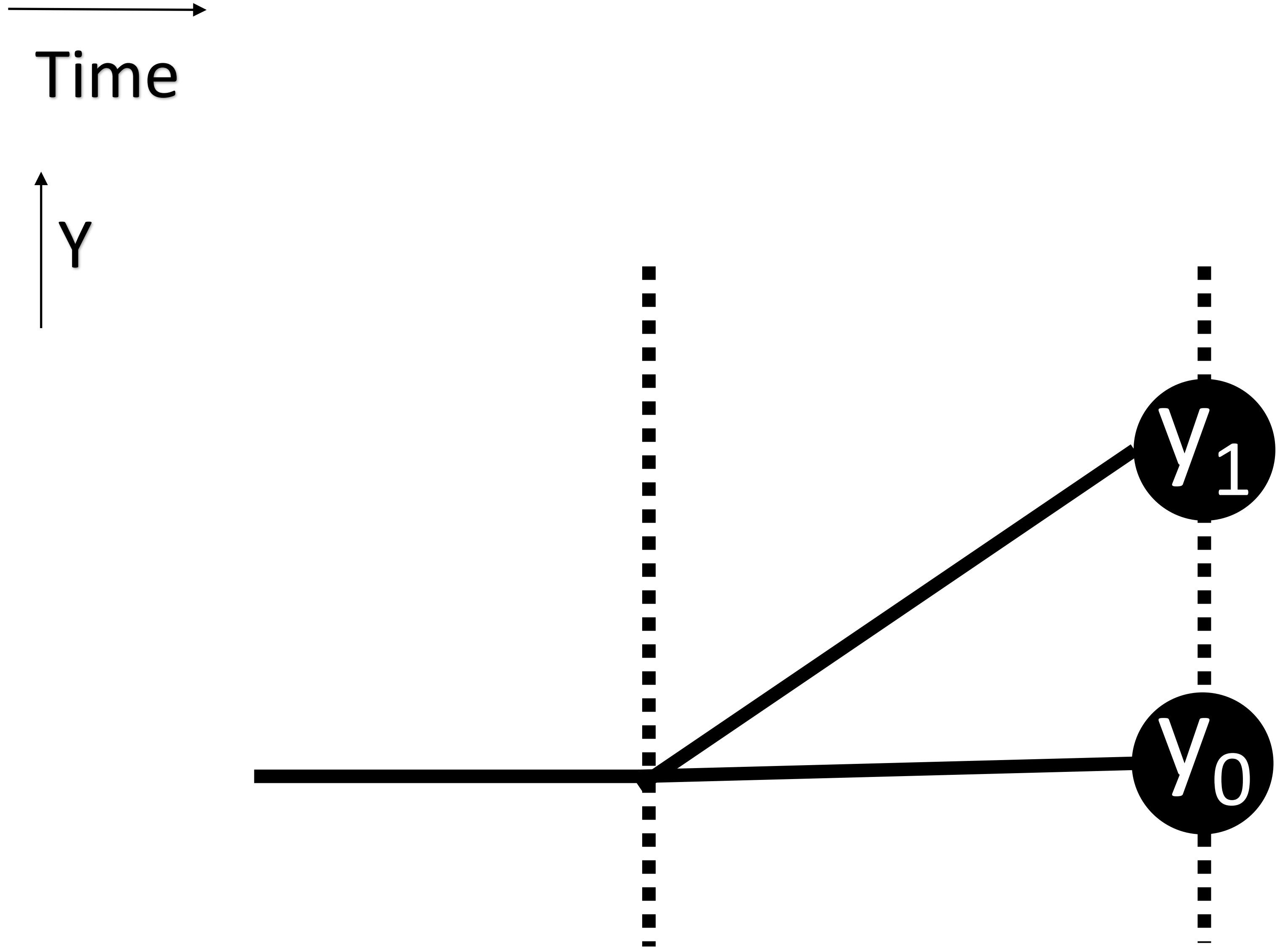
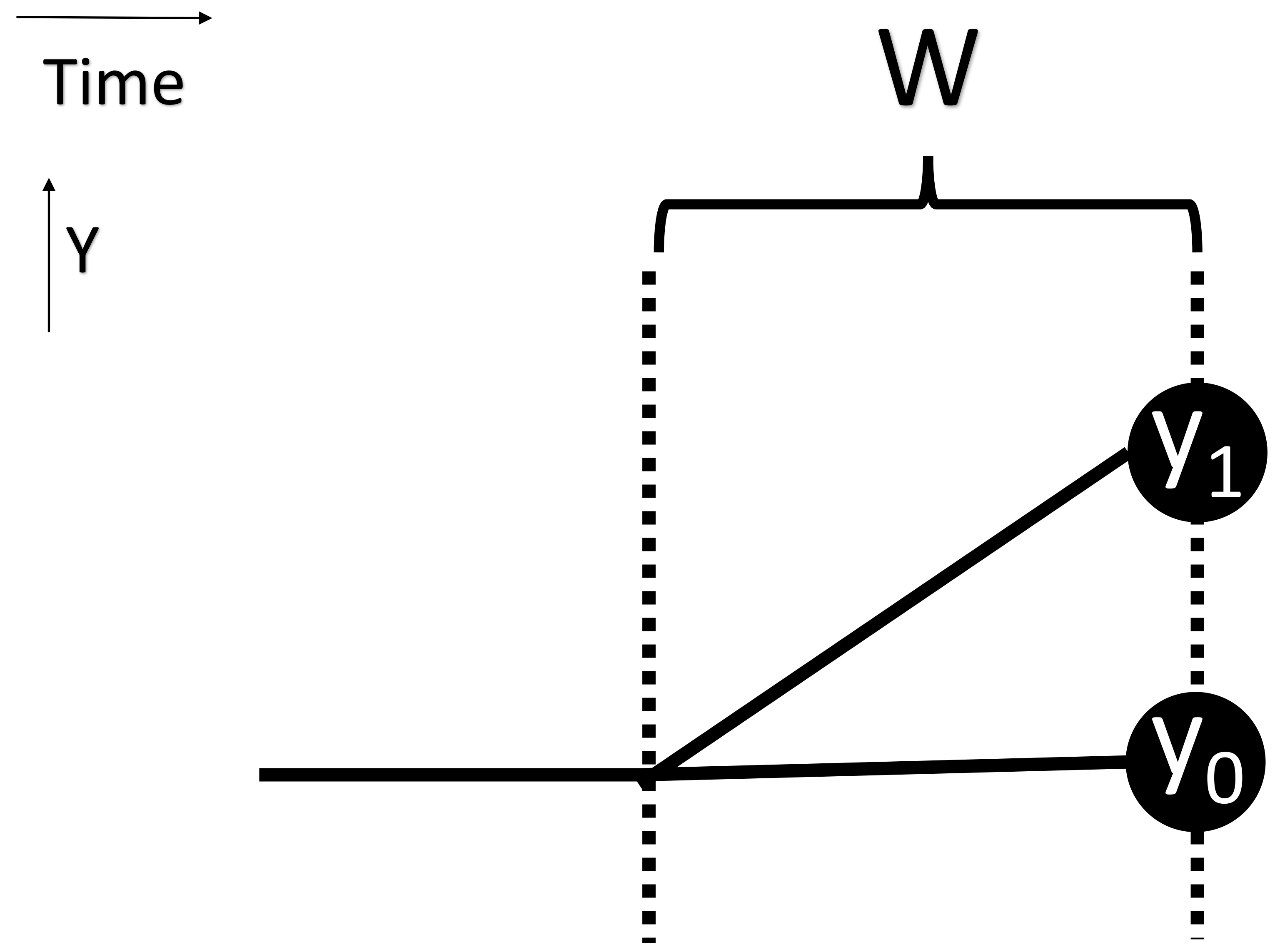
The Potential Outcome Framework

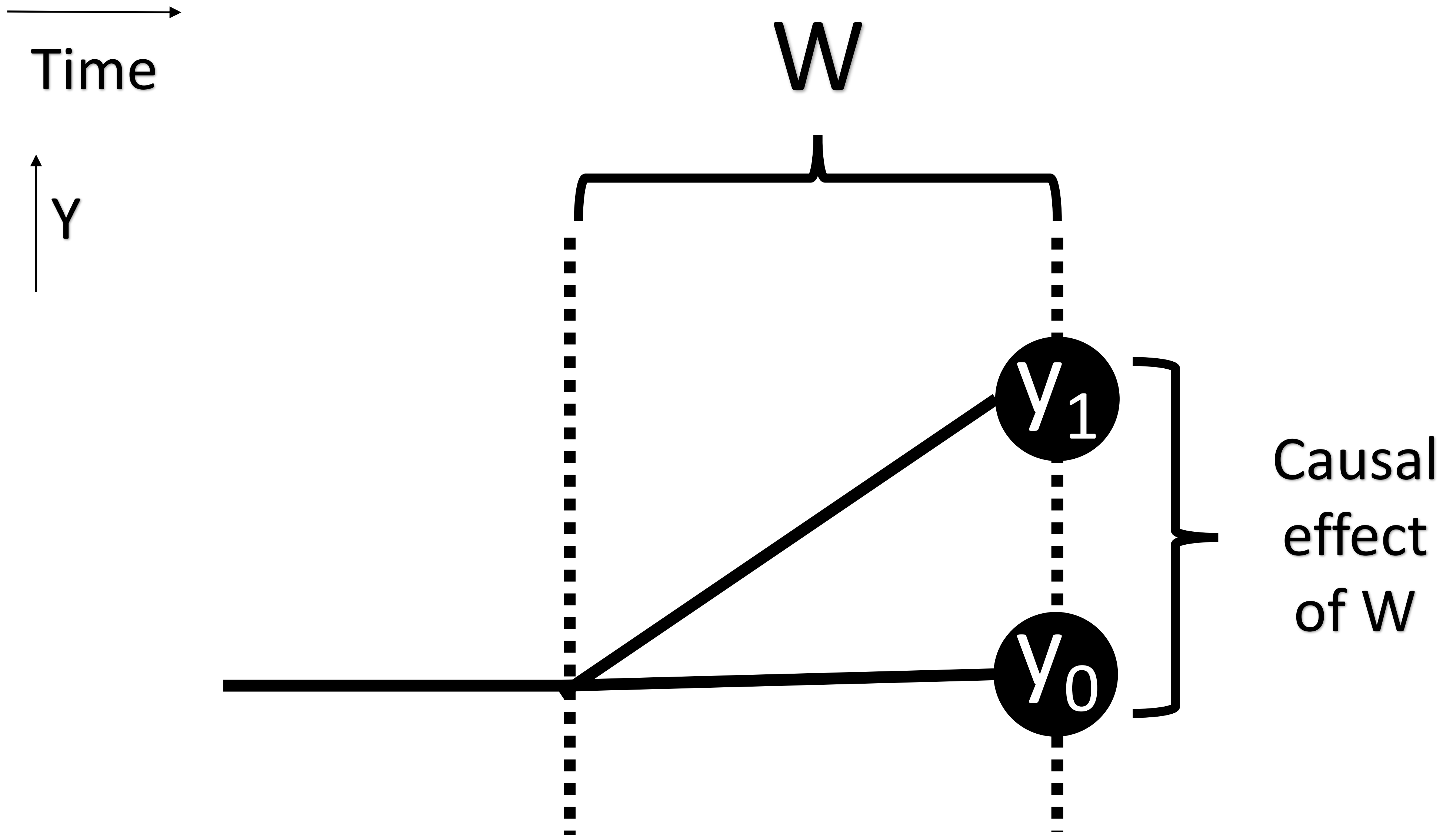
What’s in a name?

- Scientists are interested in the “causes” of things
- What do we mean by cause?
- Potential outcomes framework is an attempt to formalize this



Time
Y
y_1
y_0


Time
Y
W
y1
y0



Time
Y
W
y_1
y_0
Causal effect of W

Formalizing the ATE

The causal effect of W on individual *i* is…

$$Y_{i(w=1)} - Y_{i(w=0)}$$

Then, the average treatment effect is simply…

$$E\left(Y_{i(w=1)} - Y_{i(w=0)}\right)$$

That’s great! So, we should be able to easily estimate this, right?

"Have you ever imagined a world with no hypothetical situations?"

Estimating the ATE

So there's no way that we can measure the ATE…

$$\tau = E(Y_{i(w=1)} - Y_{i(w=0)})$$

So we're going to estimate it with the following…

$$\hat{\tau} = E\left(Y_{i(w=1)} \middle| W_i = 1\right) - E\left(Y_{i(w=0)} \middle| W_i = 0\right)$$

Ok, so this seems kind of fishy…

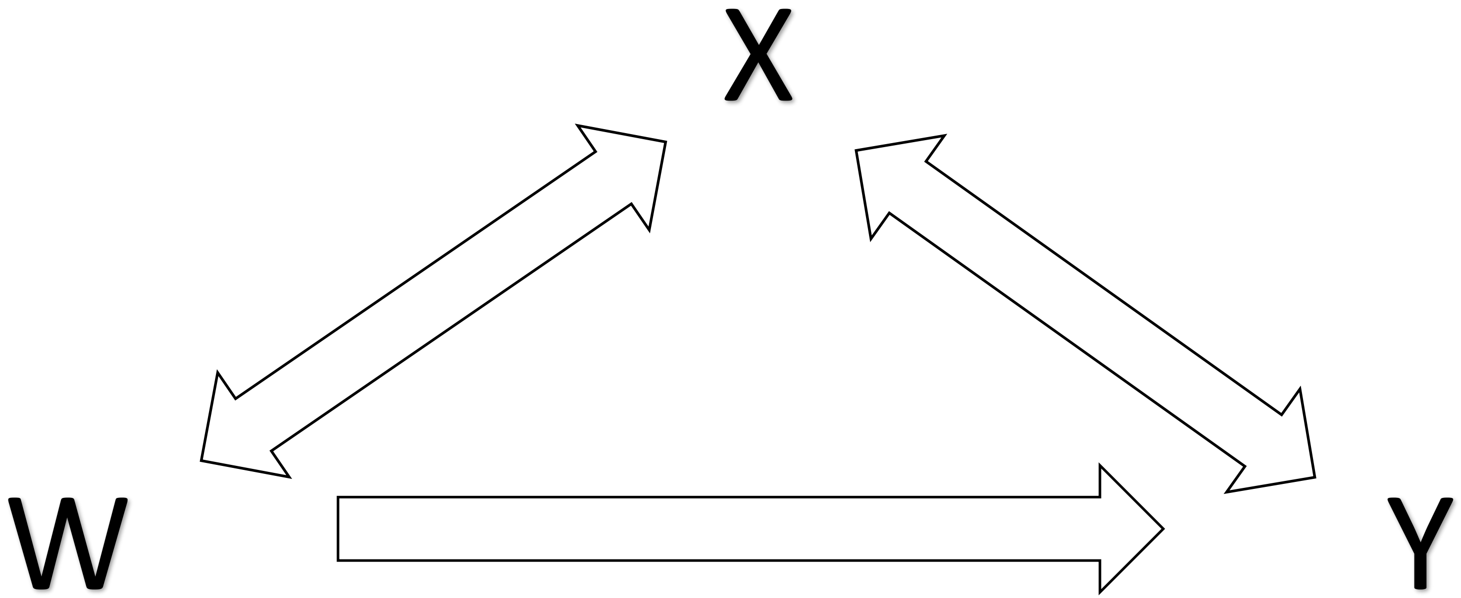
Assumptions behind ATE estimate

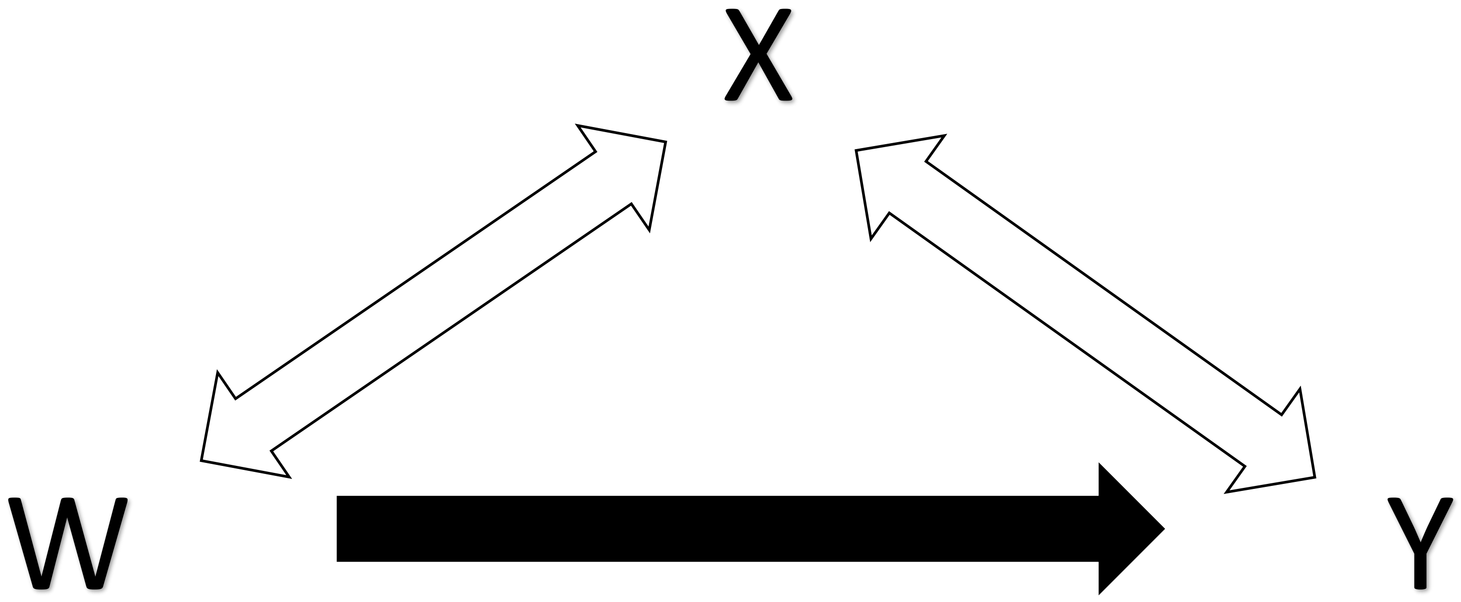
Now, it turns our there are a couple of different assumptions we could make to validate our ATE measure. Let's think about this one…

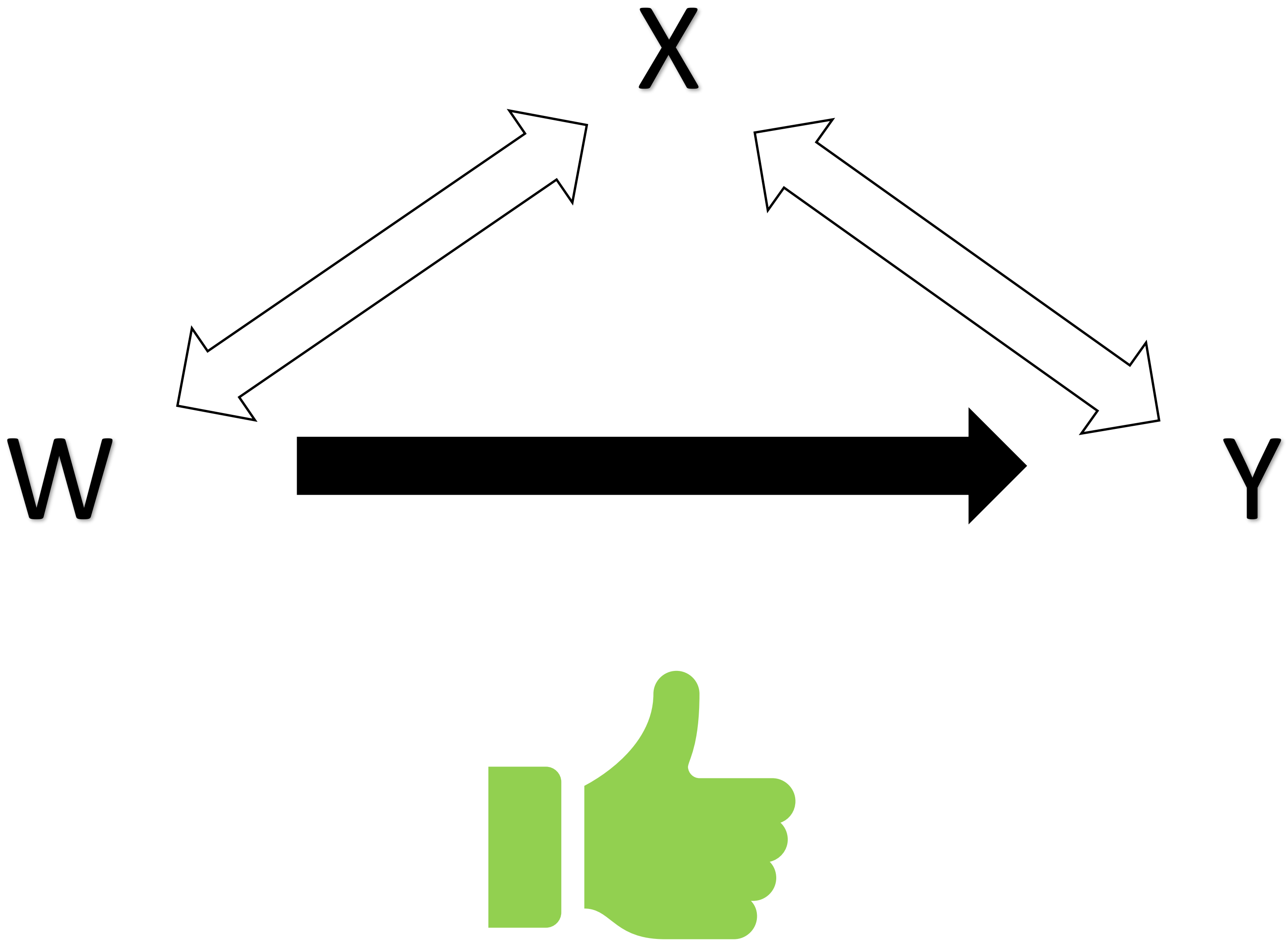
$$W_i \perp \ Y_{i(w=0)}, Y_{i(w=1)}$$

One way to get this is to have "random assignment". But how does that work?

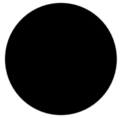
(NOTE: We also have to assume that the potential outcomes of one unit don't depend on the treatment of other units, but let's skip over this for today)


X
W
Y


X
W
Y

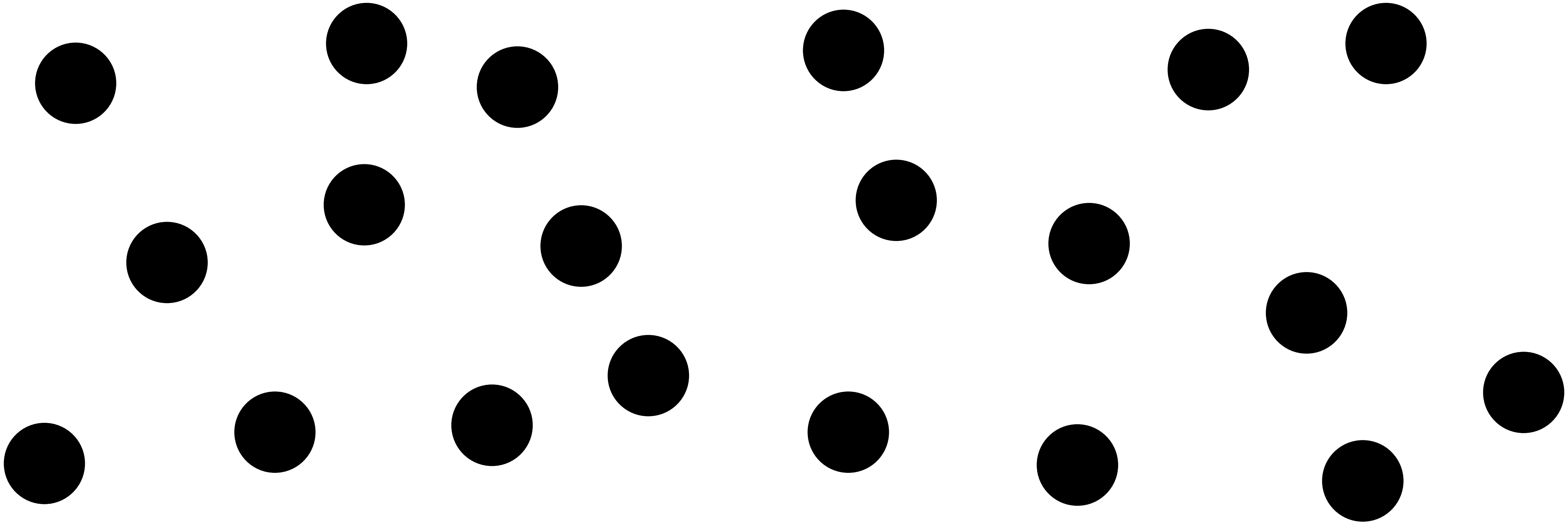


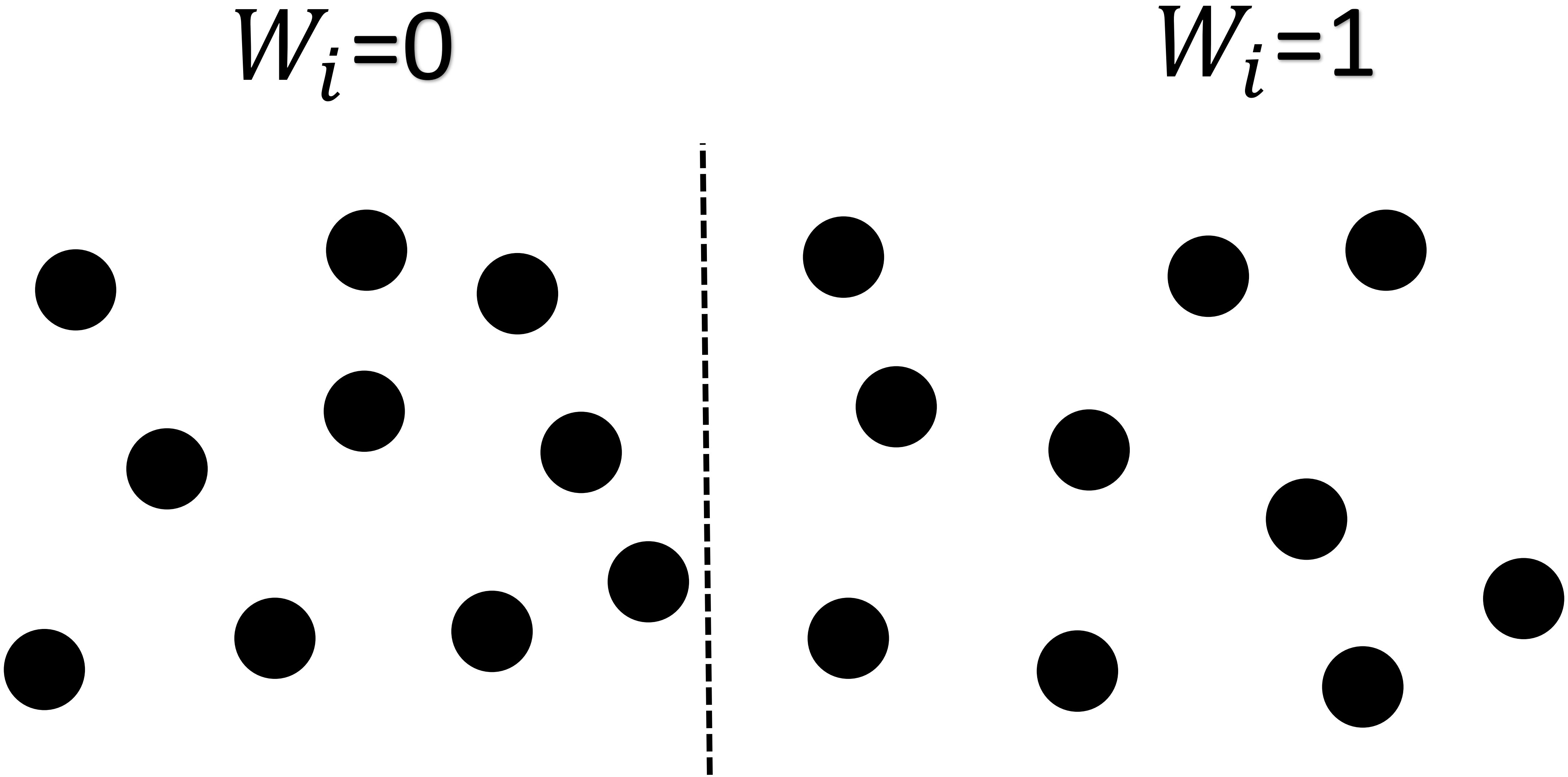
X
W
Y



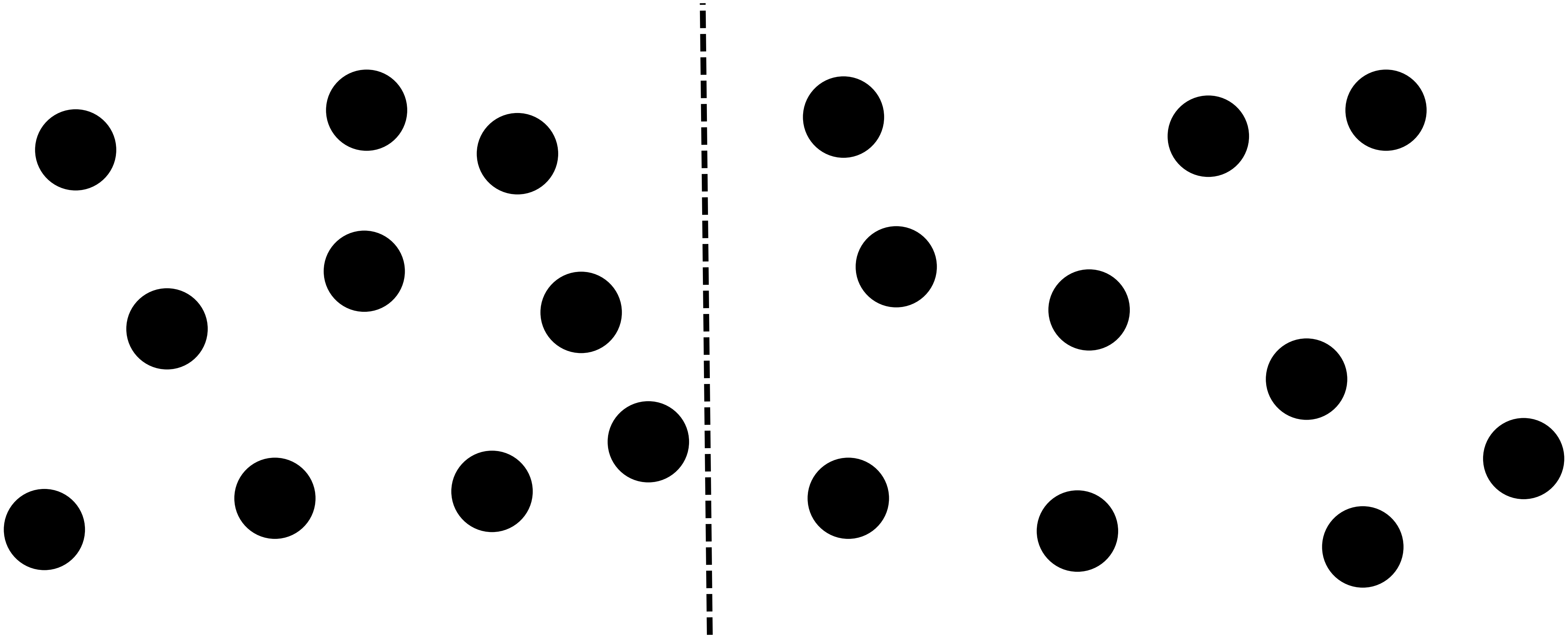
: $Y_{i(w=1)} = 10, Y_{i(w=0)} = 5$

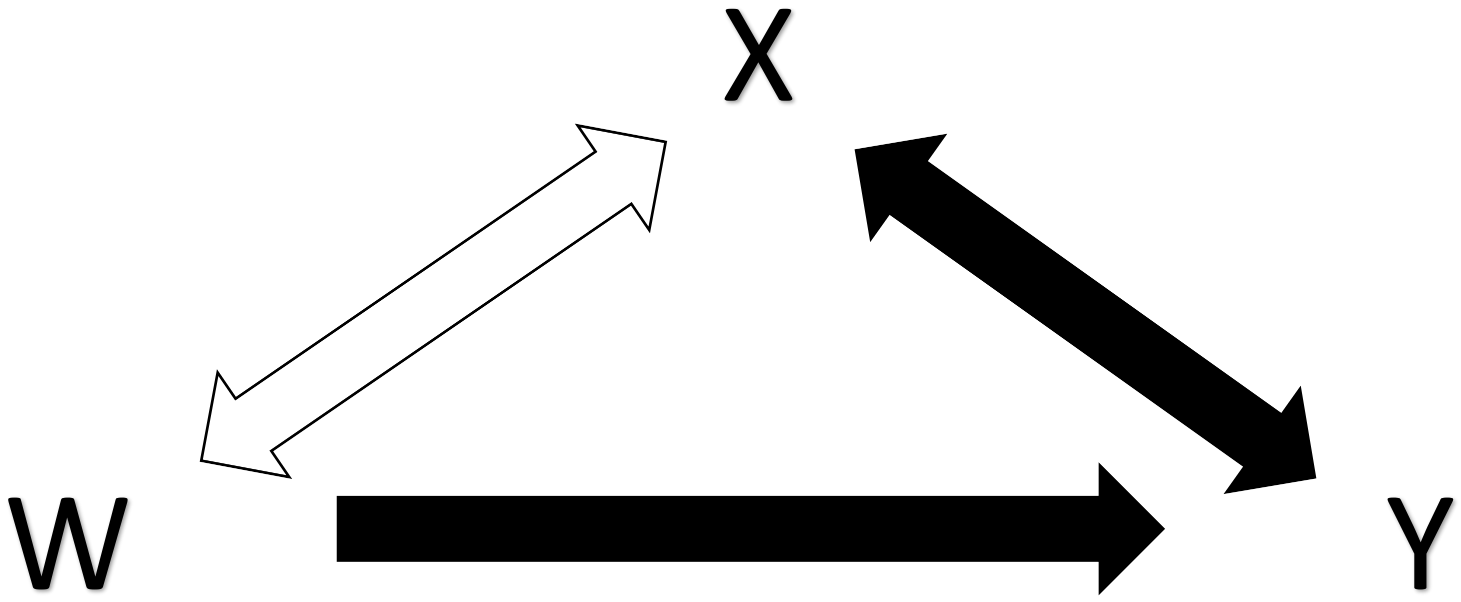
$$E\left(Y_{i(w=1)} - Y_{i(w=0)}\right) = \frac{[(10-5)*20]}{20} = 5$$

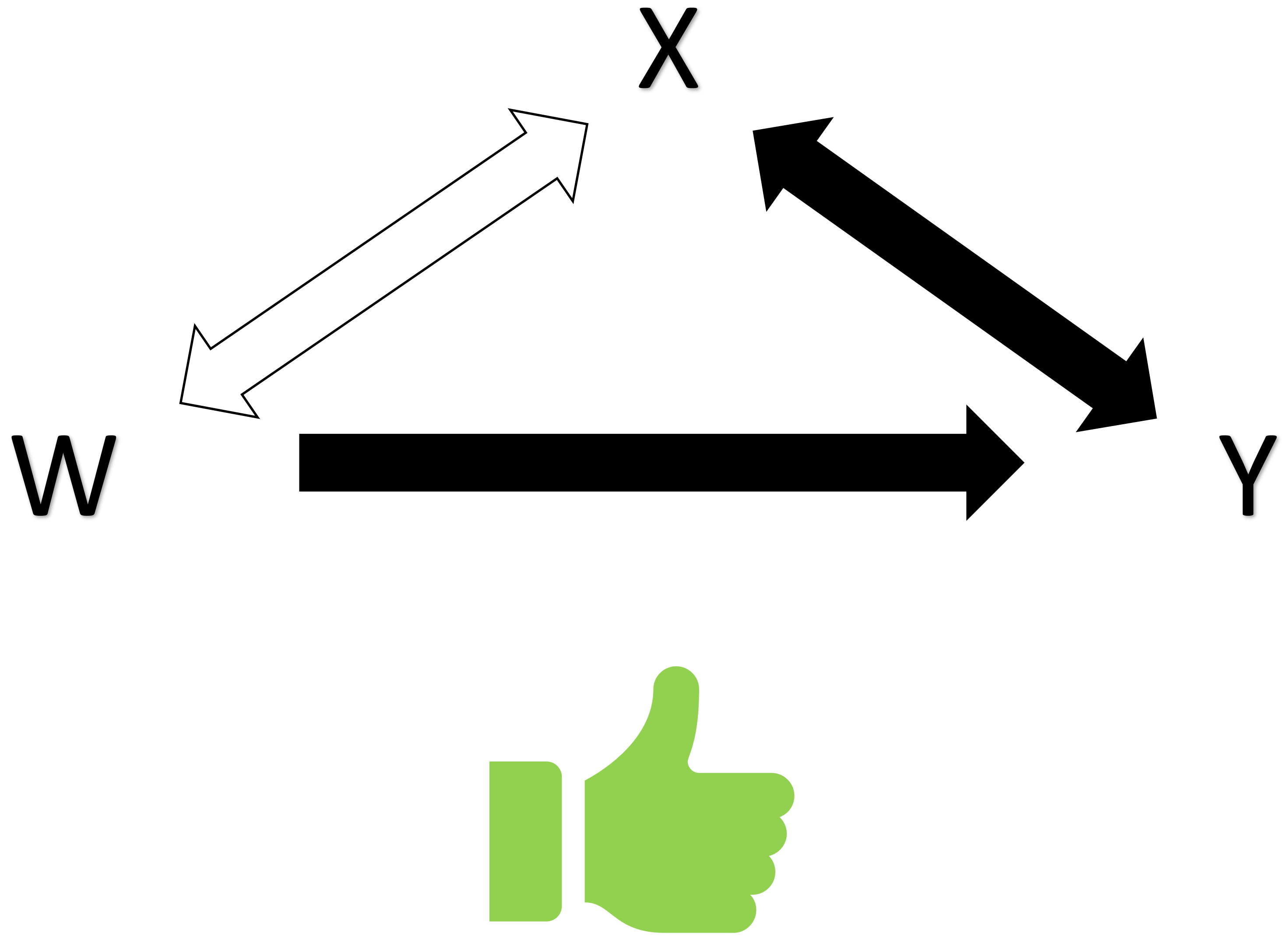



W_i=0
W_i=1

$$\hat{\tau} = E\left(Y_{i(w=1)} \middle| W_i = 1\right) - E\left(Y_{i(w=0)} \middle| W_i = 0\right) = \frac{(10\ * 10)}{10} - \frac{(5\ * 10)}{10} = 5$$



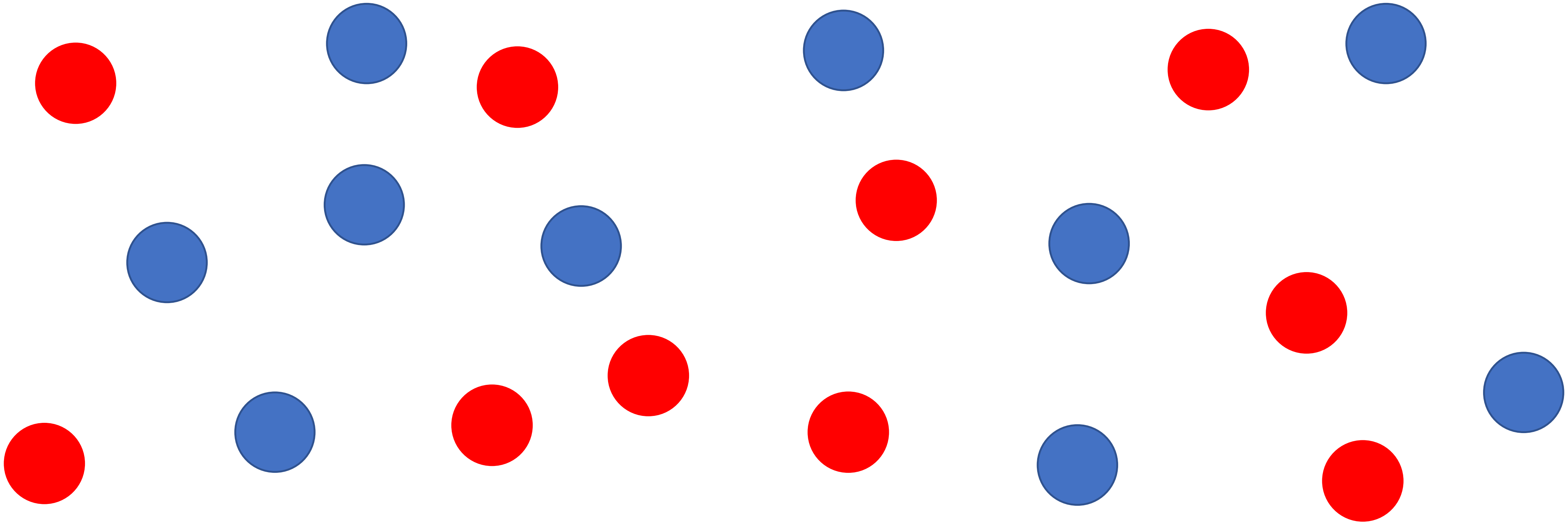

X
W
Y

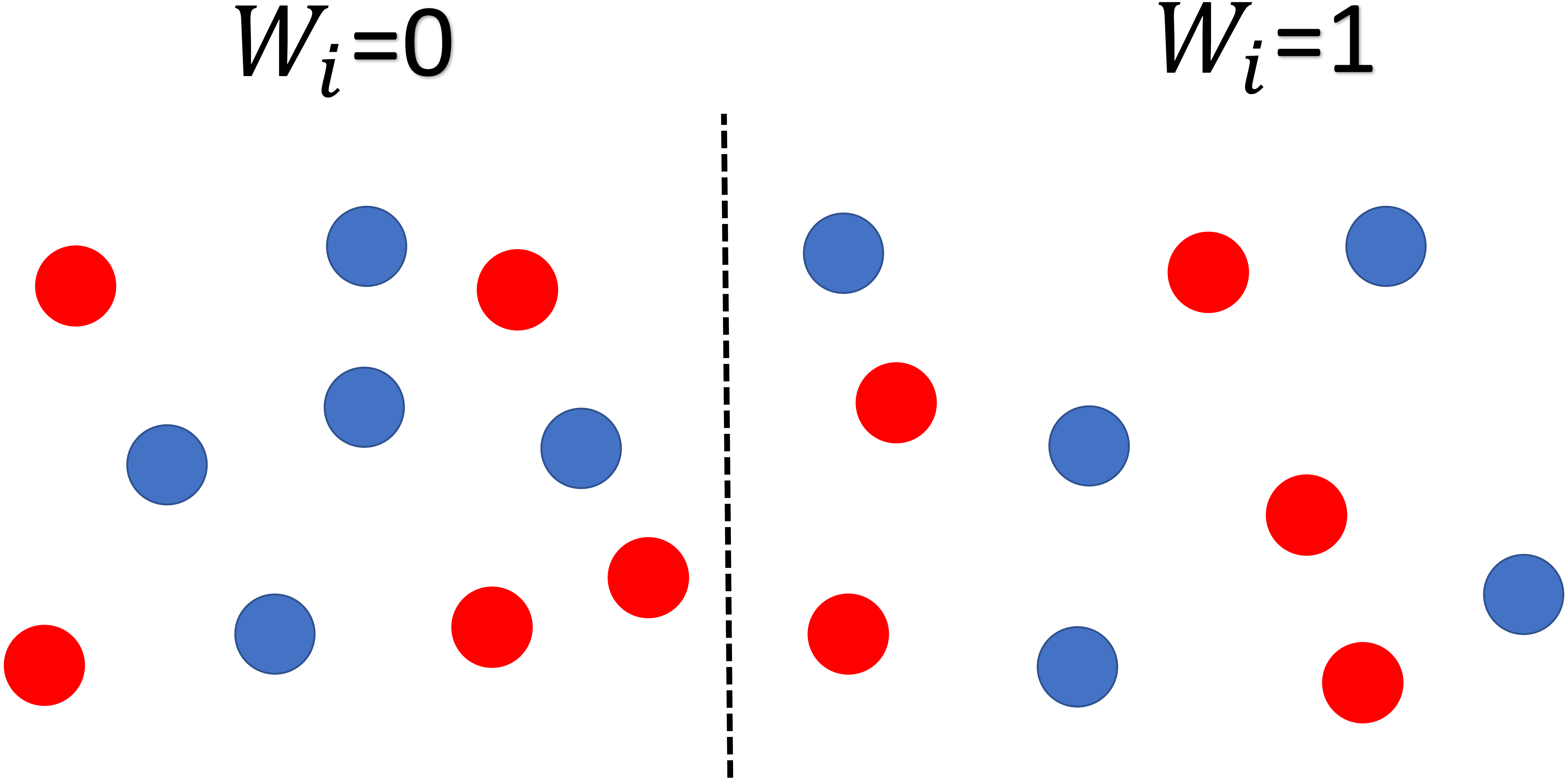

X
W
Y

$: Y_{i(w=1)} = 10, Y_{i(w=0)} = 5$

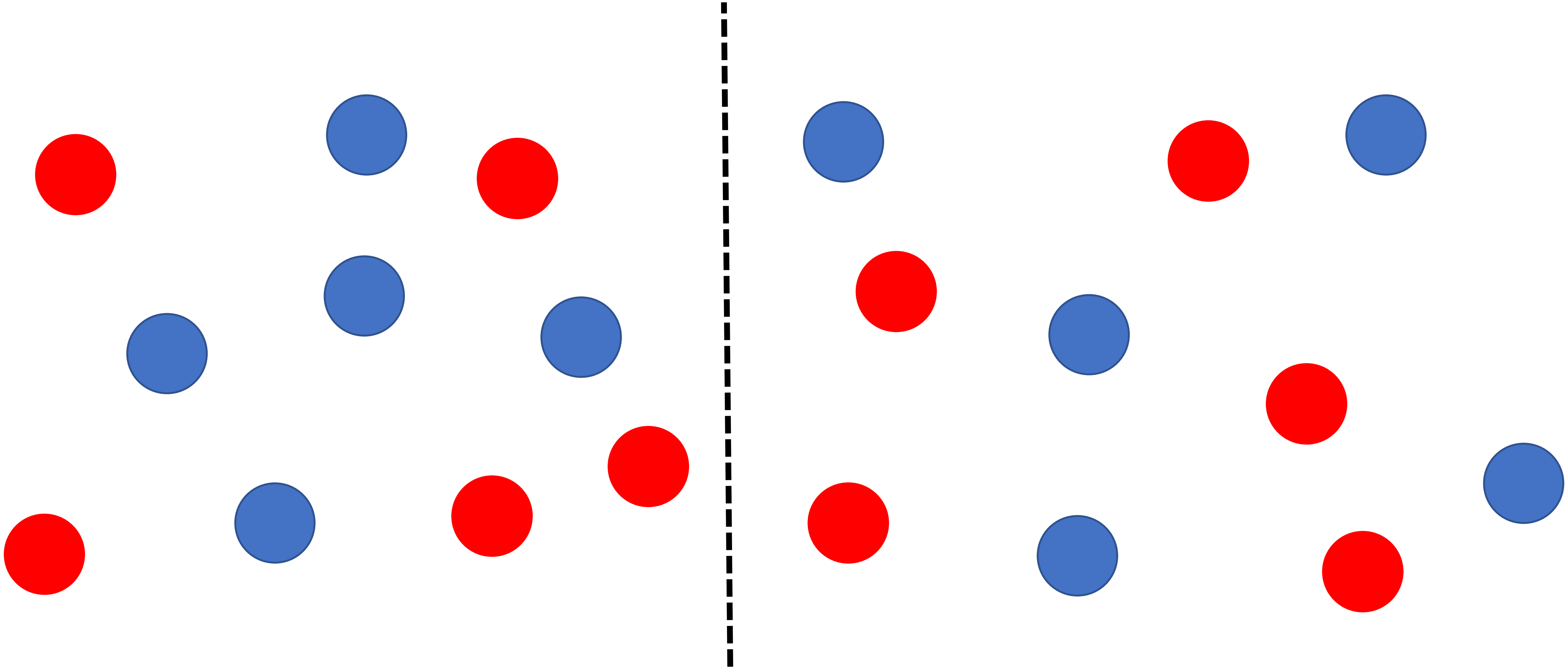
$: Y_{i(w=1)} = 8, Y_{i(w=0)} = 5$

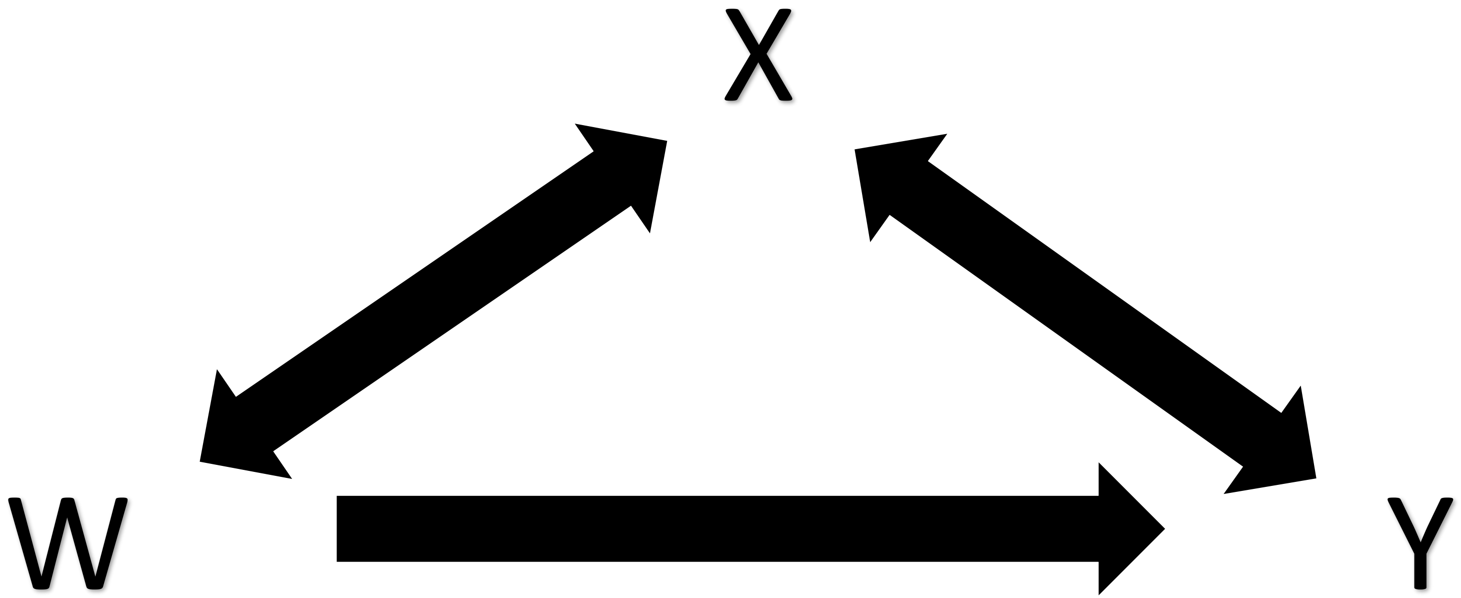
$$E(Y_{i(w=1)} - Y_{i(w=0)}) = \frac{[(10-5)*10]+[(8-5)*10]}{20} = 4$$

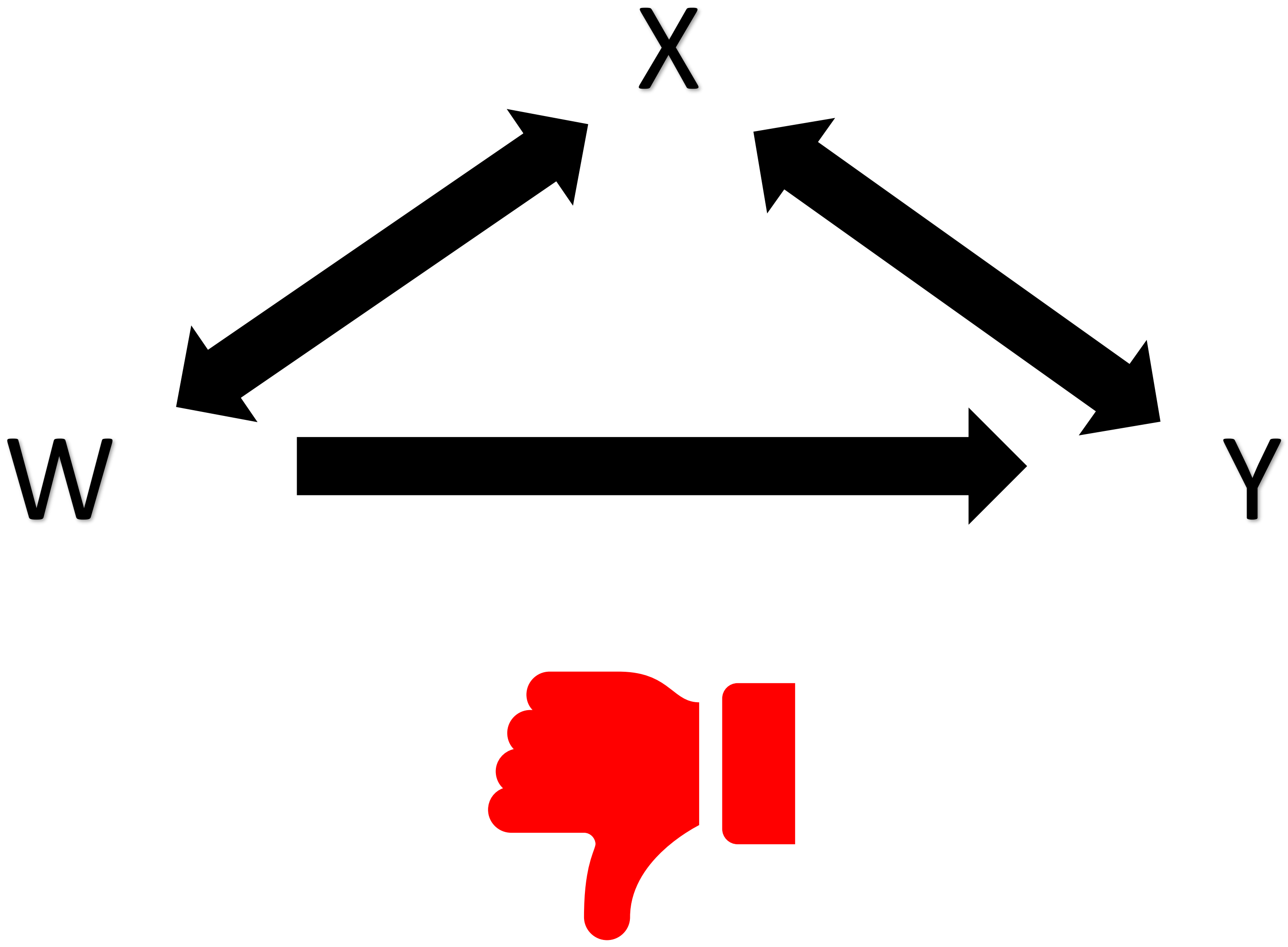


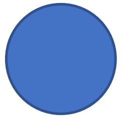

W_i=0
W_i=1

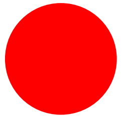
$$\hat{\tau} = E\left(Y_{i(w=1)} \middle| W_i = 1\right) - E\left(Y_{i(w=0)} \middle| W_i = 0\right) = \frac{(10 \ * 5) + (8 * 5)}{10} - \frac{(5 \ * 5) + (5 * 5)}{10} = 4$$



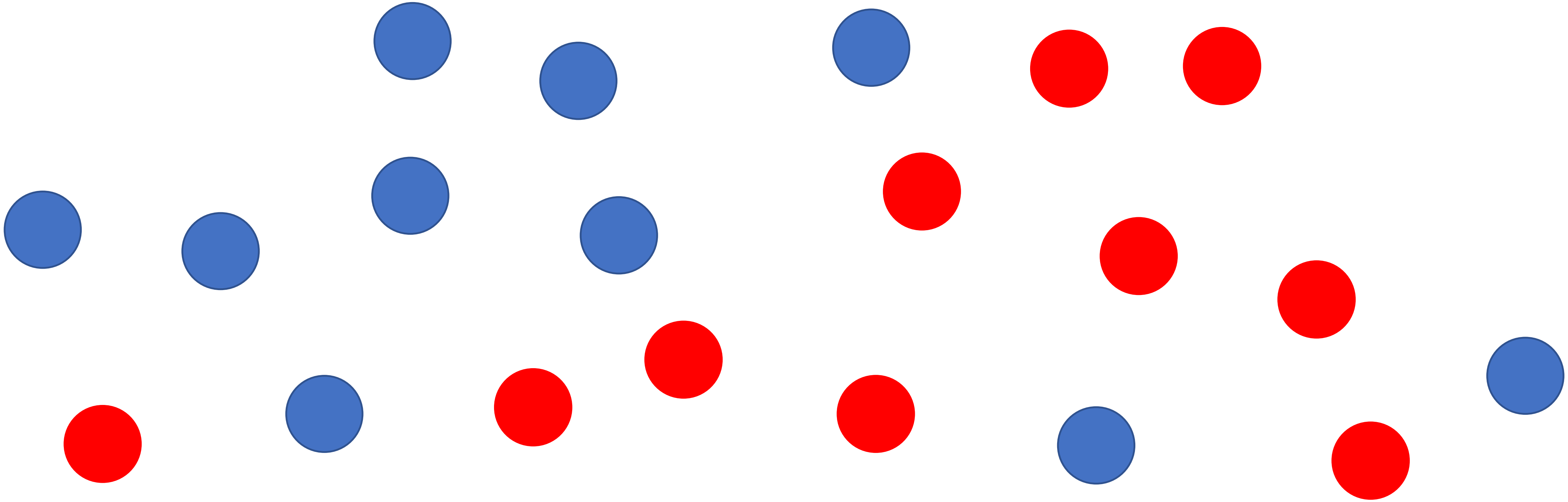

X
W
Y

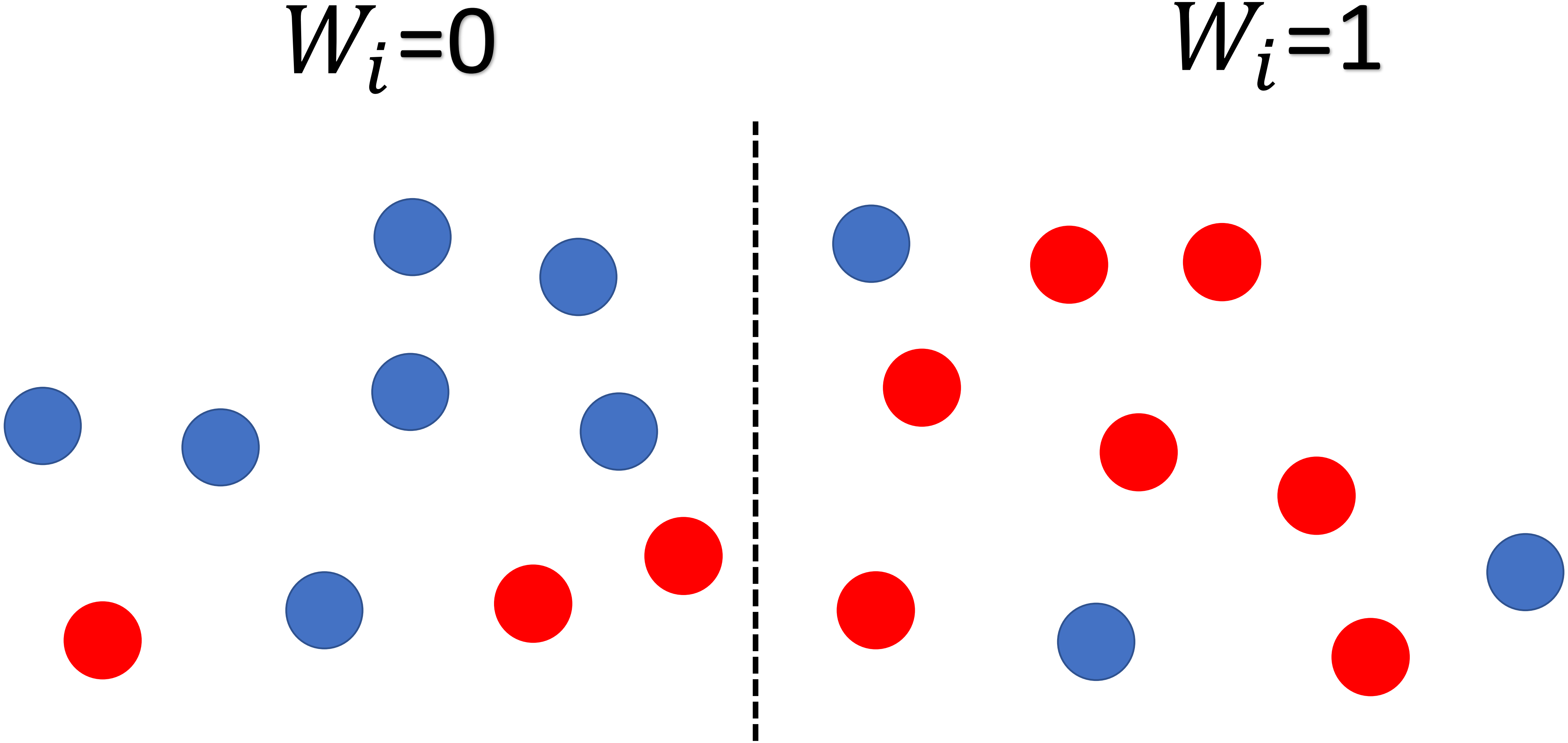

X
W
Y

 $: Y_{i(w=1)} = 10, Y_{i(w=0)} = 5$

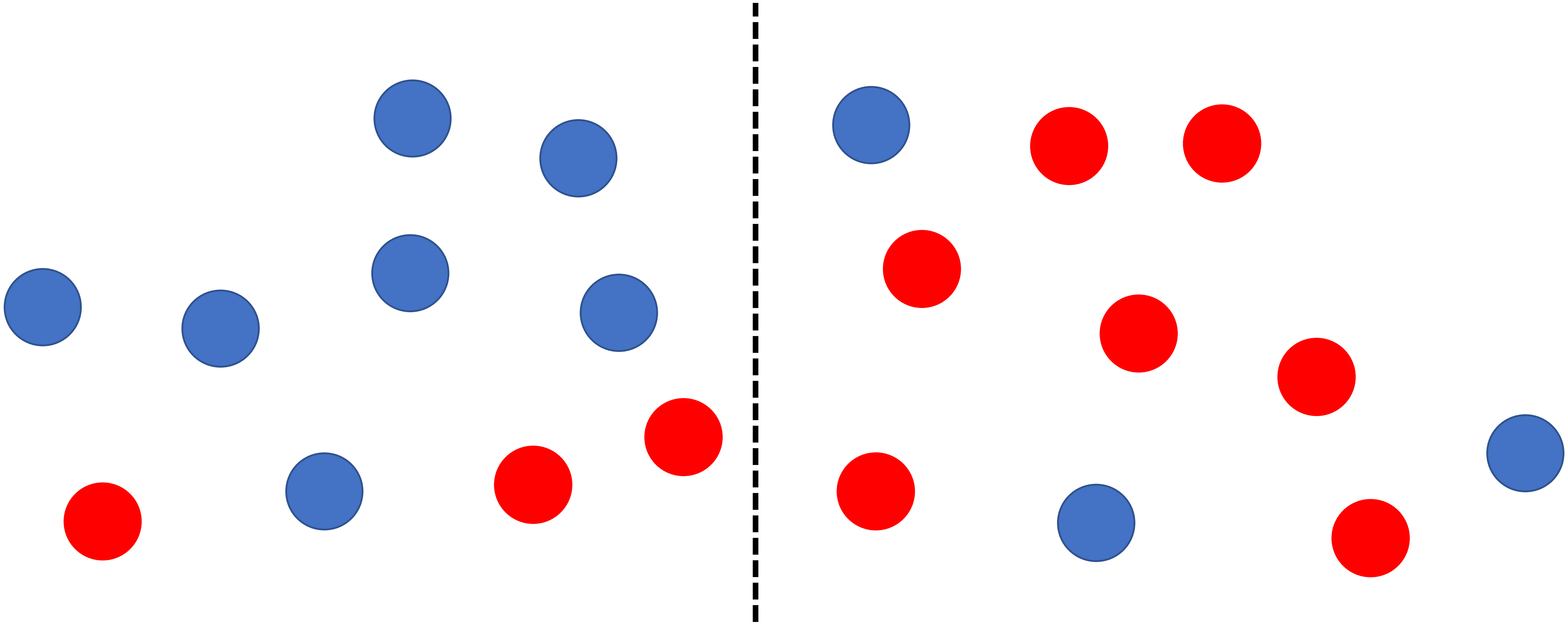
 $: Y_{i(w=1)} = 8, Y_{i(w=0)} = 5$

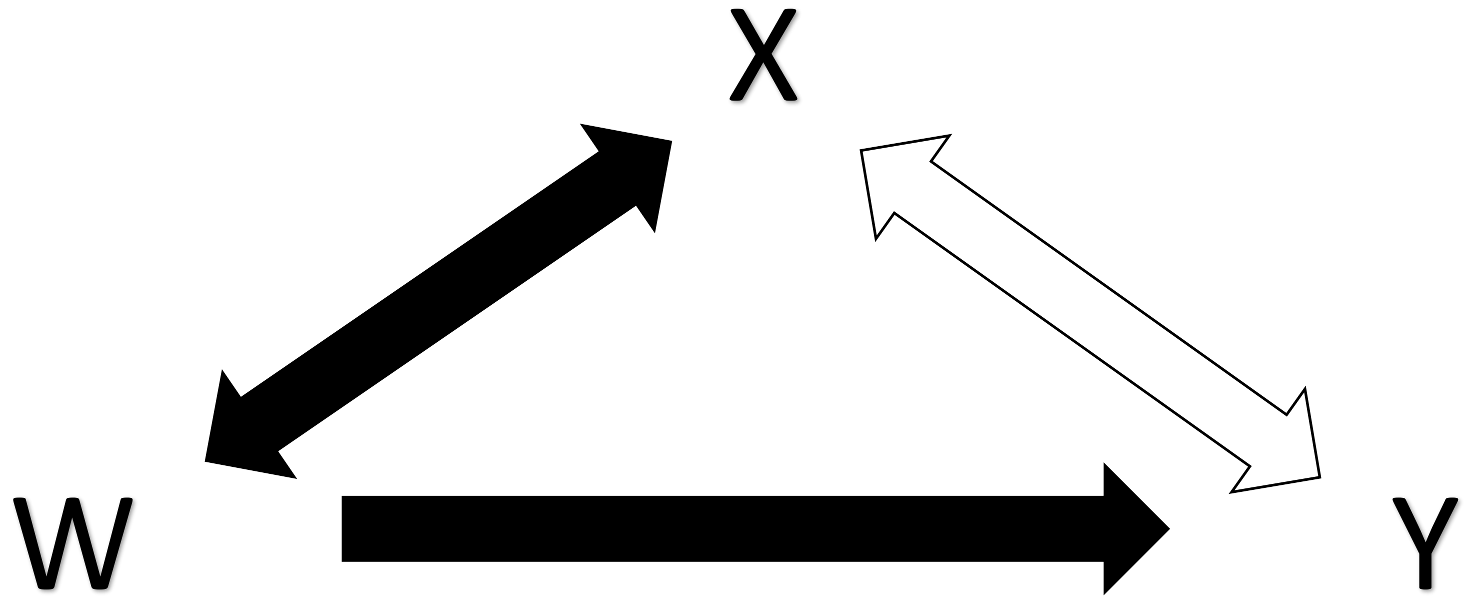
$$E(Y_{i(w=1)} - Y_{i(w=0)}) = \frac{[(10-5)*10]+[(8-5)*10]}{20} = 4$$



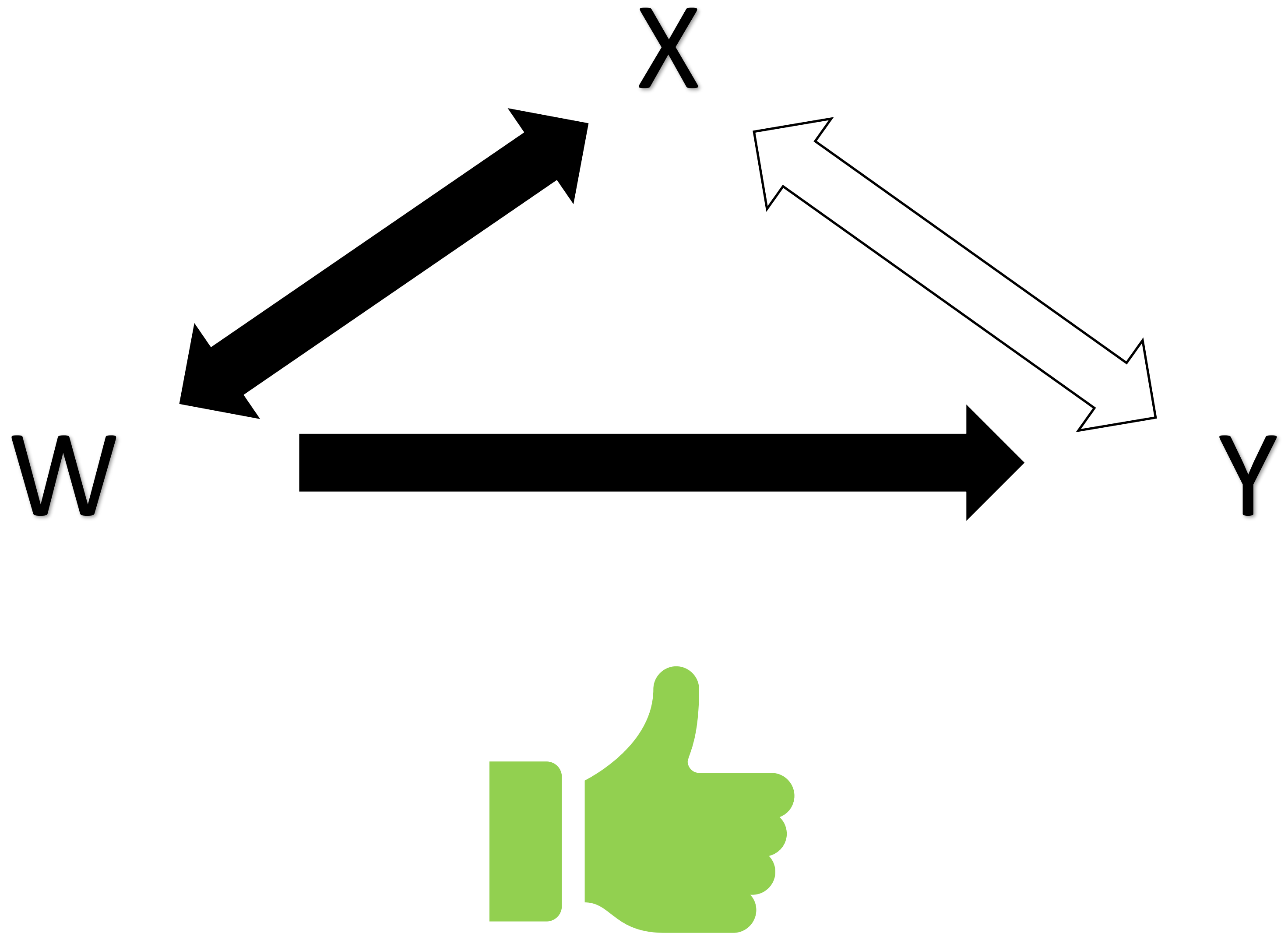

W_i=0
W_i=1

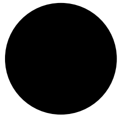
$$\hat{\tau} = E\left(Y_{i(w=1)} \middle| W_i = 1\right) - E\left(Y_{i(w=0)} \middle| W_i = 0\right) = \frac{(10 \ * 7) + (8 * 3)}{10} - \frac{(5 \ * 3) + (5 * 7)}{10} = 4.4$$





X
W
Y

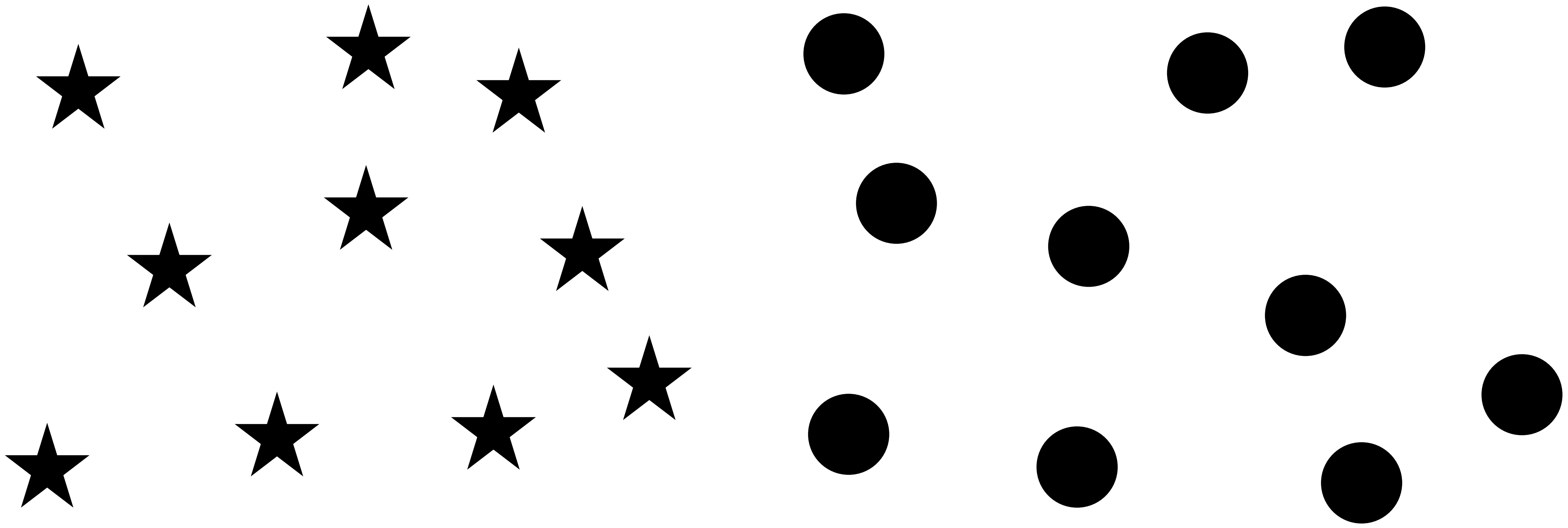

X
W
Y

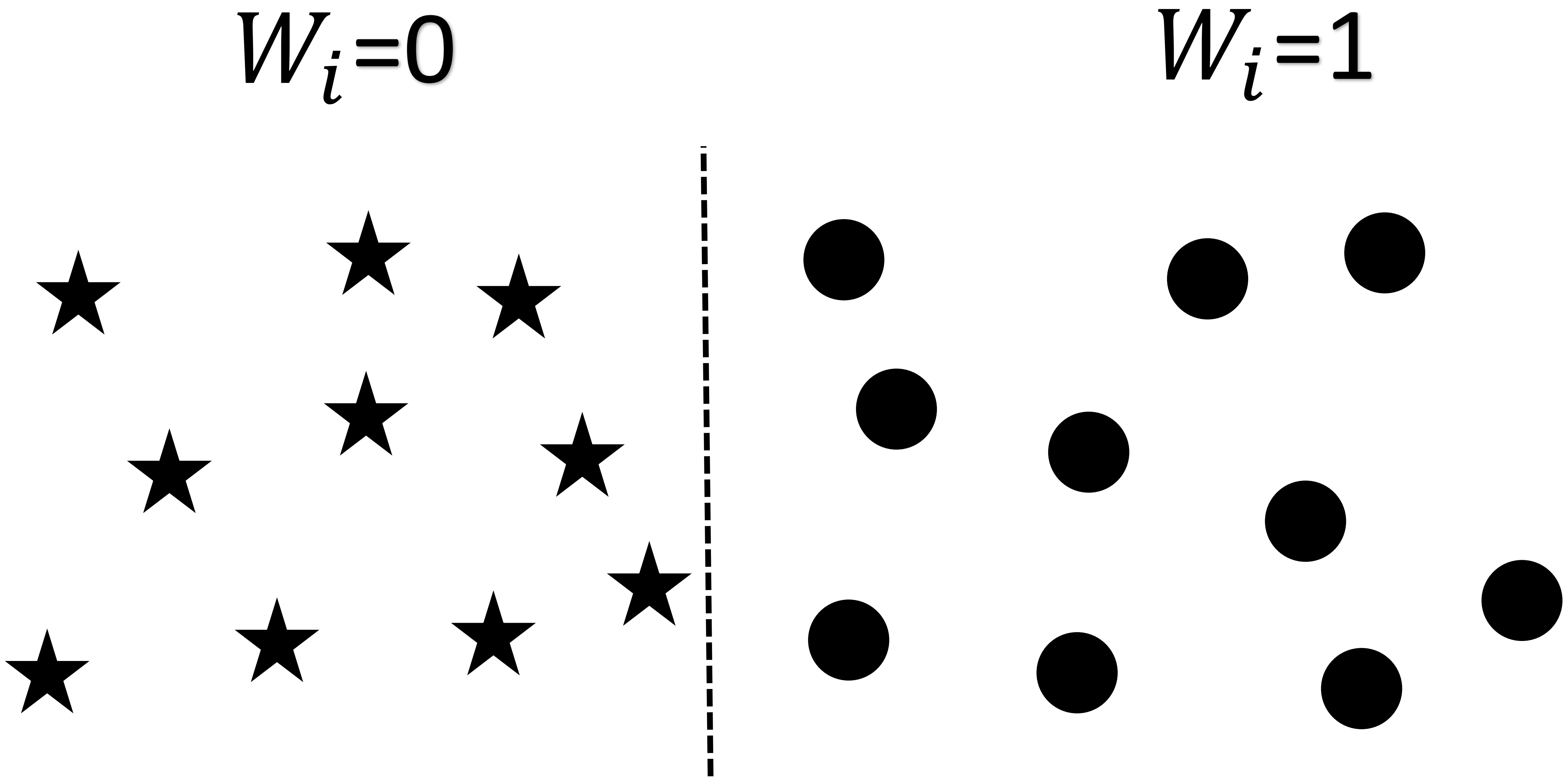


$: Y_{i(w=1)} = 10, Y_{i(w=0)} = 5$

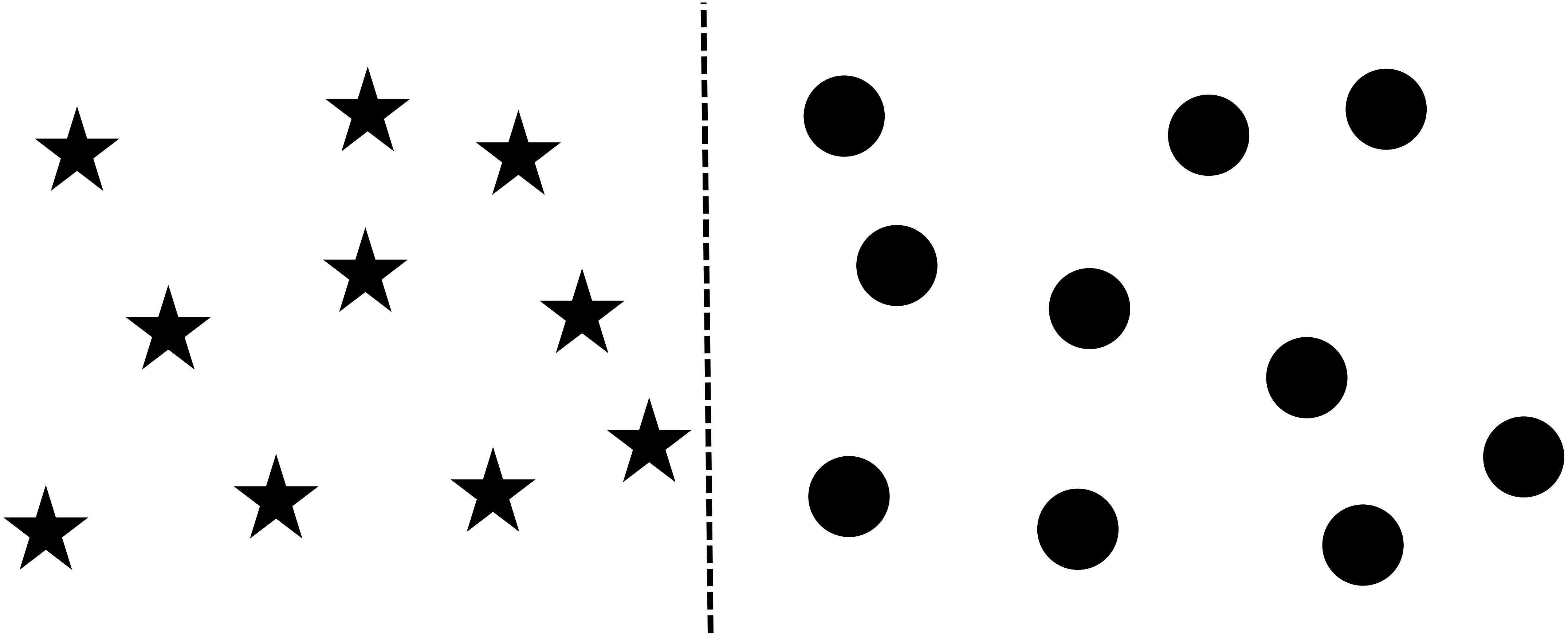
$: Y_{i(w=1)} = 10, Y_{i(w=0)} = 5$

$$E\left(Y_{i(w=1)} - Y_{i(w=0)}\right) = \frac{[(10-5)*20]}{20} = 5$$




W_i=0
W_i=1

$$\hat{\tau} = E\left(Y_{i(w=1)} \middle| W_i = 1\right) - E\left(Y_{i(w=0)} \middle| W_i = 0\right) = \frac{(10\ * 10)}{10} - \frac{(5\ * 10)}{10} = 5$$



Thinking about treatments

- What kinds of things can be treatments? What can’t?
- Can we “treat” groups? Organizations? Social networks? Countries?
- What are the limitations of controlled laboratory experiments?

Observational studies

But, sometimes it's impossible (or unethical) to randomize! Let's re-examine our assumption…

$$W_i \perp Y_{i(w=0)}, Y_{i(w=1)}$$

This might not be reasonable to assume in observational studies. It turns out that we can make the slightly more relaxed assumption…

$$W_i \perp Y_{i(w=0)}, Y_{i(w=1)} | X_i$$

Basically, if the only thing that is biasing our treatment are **observables**, then we can correct, if we make the ATE estimate a little more complicated…

$$\hat{\tau} = E[E\left(Y_{i(w=1)} \middle| W_i = 1\right) - E\left(Y_{i(w=0)} \middle| W_i = 0\right) | X_i]$$

This means that if our conditions are perfectly matched on all observables, then we can recover the ATE!

Observational studies (continued)

You can derive an even nicer assumption…

$$W_i \perp Y_{i(w=0)}, Y_{i(w=1)} | e_i$$

Where e_i is…

$$e_i = P(W_i = 1 | X_i)$$

If you're again willing to make the estimating equation for the ATE a bit more complicated…

$$\hat{\tau} = E(\frac{W_i Y_i}{\hat{e_i}} - \frac{(1-W_i) Y_i}{1 - \hat{e_i}})$$

Now, I know this equation looks scary, but let's walk through it…

$$\hat{\tau} = E\left(\frac{W_i Y_i}{\widehat{e_i}} - \frac{(1-W_i)Y_i}{1-\widehat{e_i}}\right)$$

When $W_i = 1$…

$$\hat{\tau} = E\left(\frac{W_i Y_i}{\hat{e}_i} - \cancel{\frac{(1-W_i)Y_i}{1-\hat{e}_i}}\right)$$

When $W_i = 0$…

$$\hat{\tau} = E\left(\cancel{\frac{W_i Y_i}{\hat{e}_i}} - \frac{(1-W_i)Y_i}{1-\hat{e}_i}\right)$$

Intuition behind IPW

- When comparing individuals who have equal propensity score, treatment is “as-good-as” random
- Individuals who were (not) likely to be treated and were (not) treated are not very informative, so down-weight them
- Individuals who were very likely to be treated but were treated or who were very unlikely to be treated but were are very informative, so up-weight them