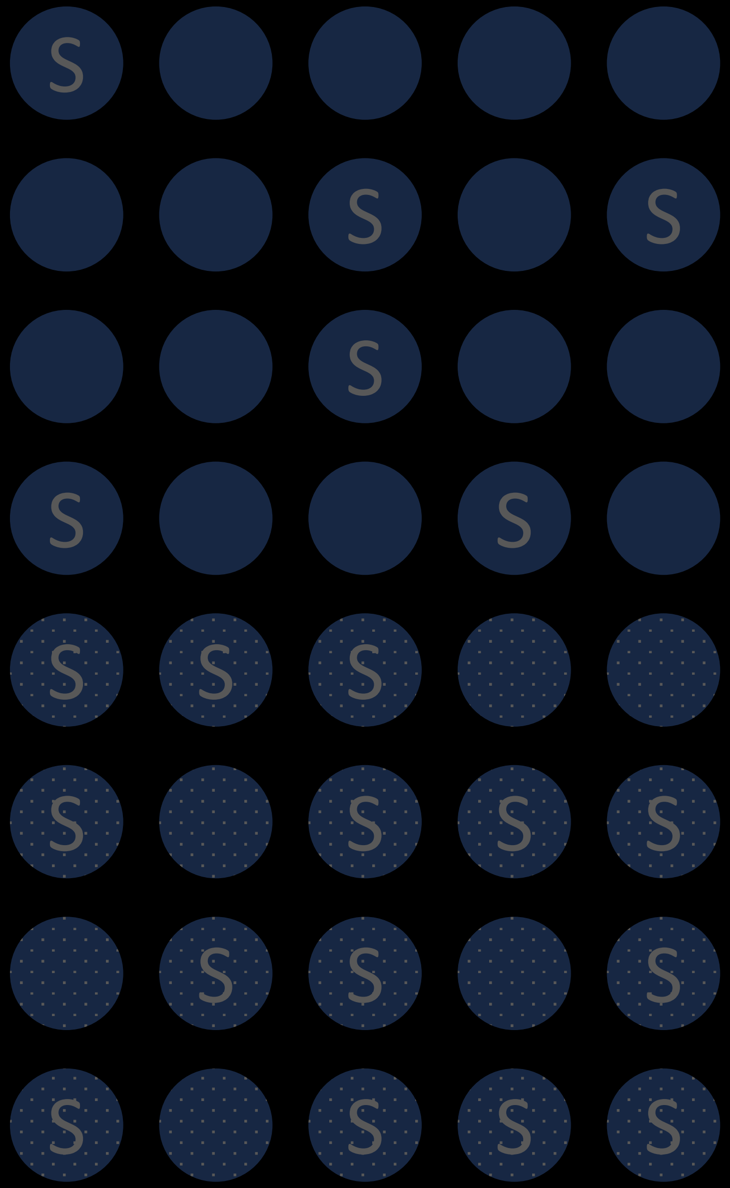
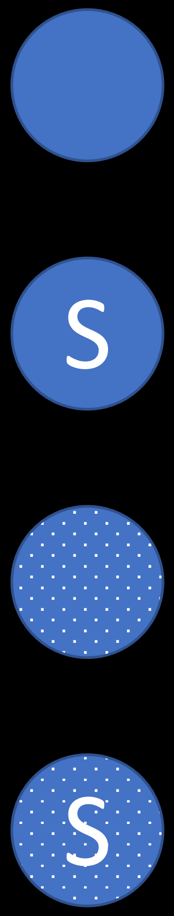
Weighting and Marginal Structural Models

Keletso Makofane
PhD Candidate in Social Network Epidemiology
Harvard TH Chan School of Public Health
June 2020

keletso.makofane@gmail.com
Twitter: @klts0





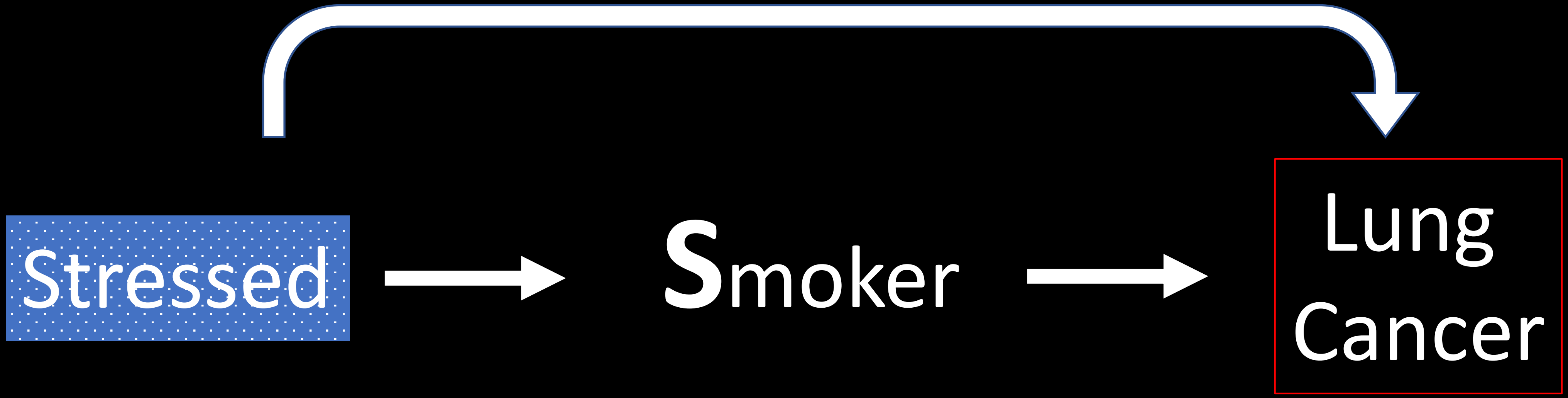
Unstressed Non-Smoker

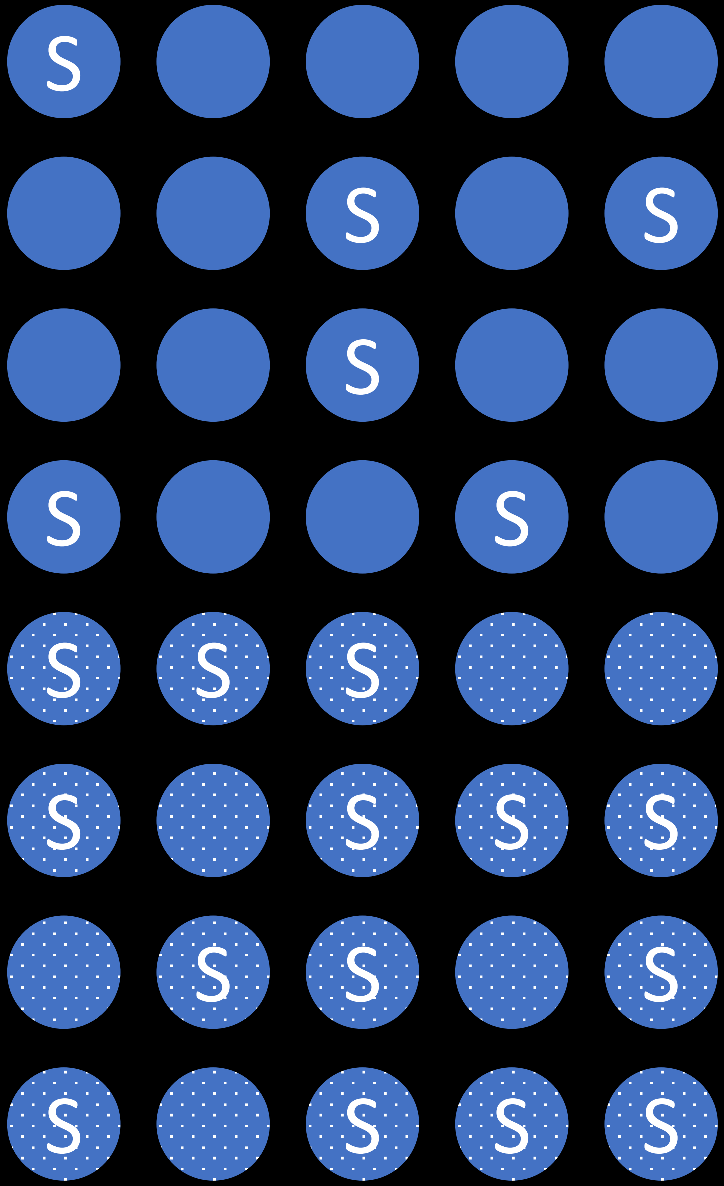
Unstressed Smoker

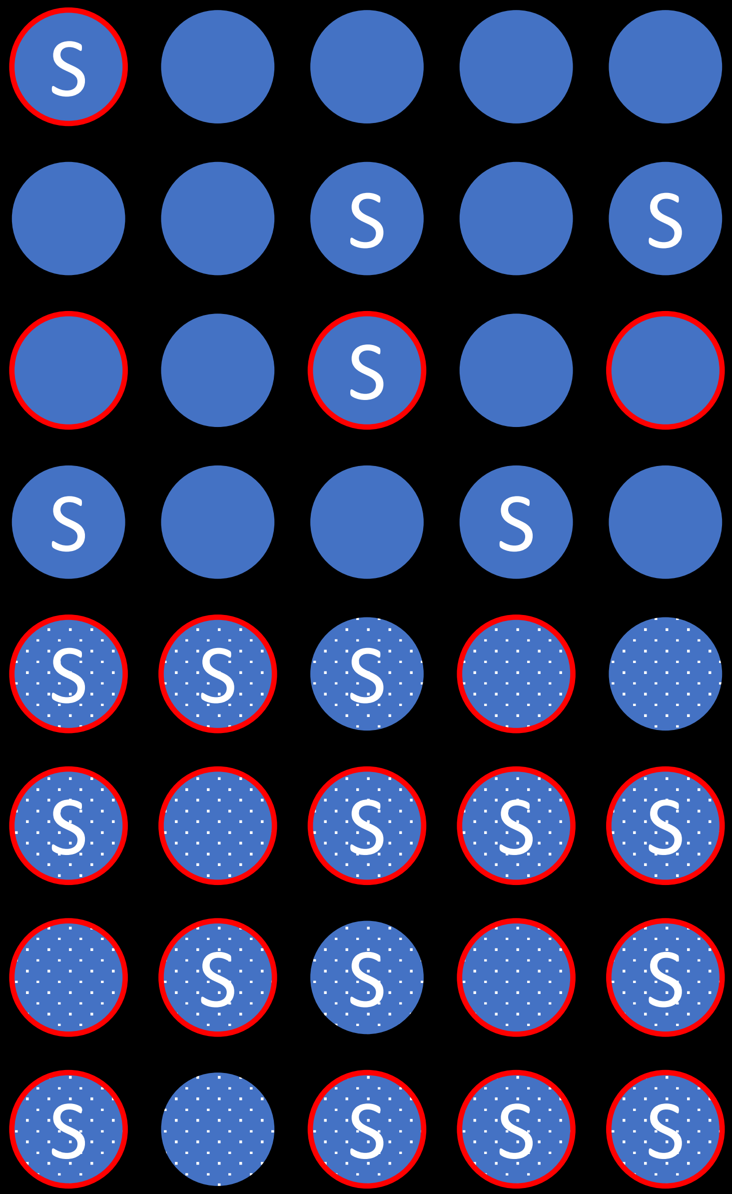
Stressed Non-Smoker

Stressed Smoker

We are interested in the effect of smoking on lung cancer



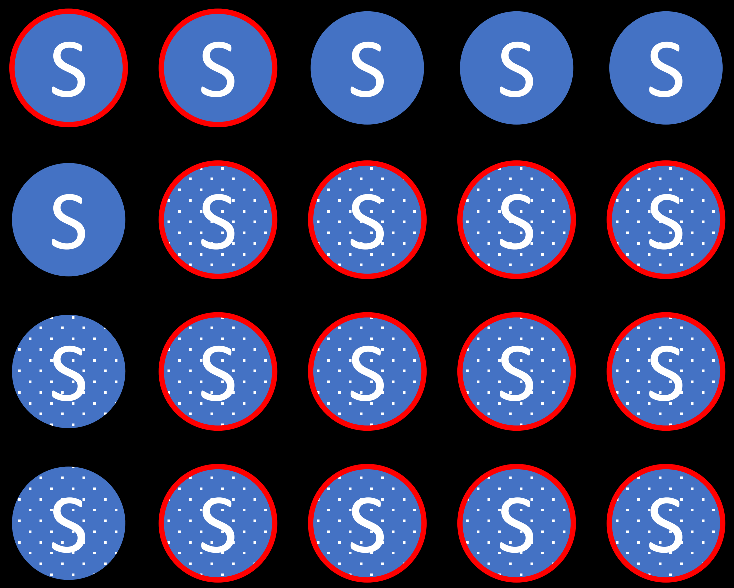

S
S S
S
S S
S S S
S S S S
S S S
S S S S



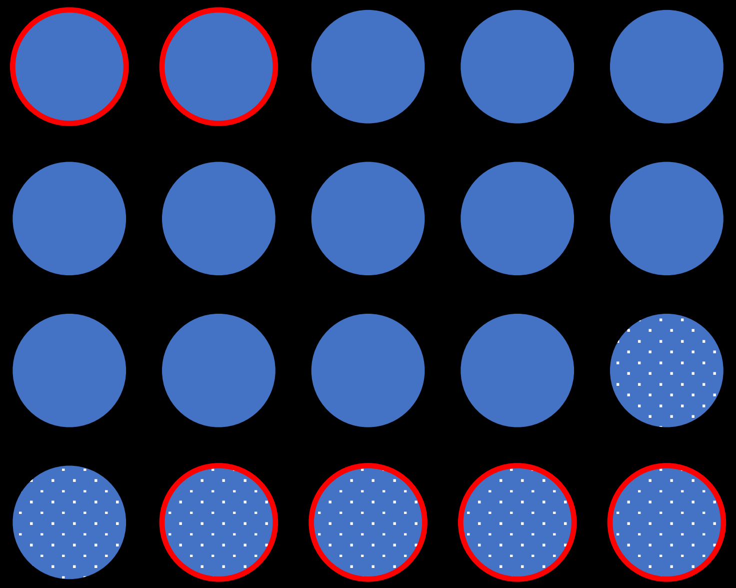
Keletso Makofane, June 2020

The effect of smoking on lung cancer

$$\frac{14}{20} - \frac{6}{20} = \frac{8}{20}\,?$$

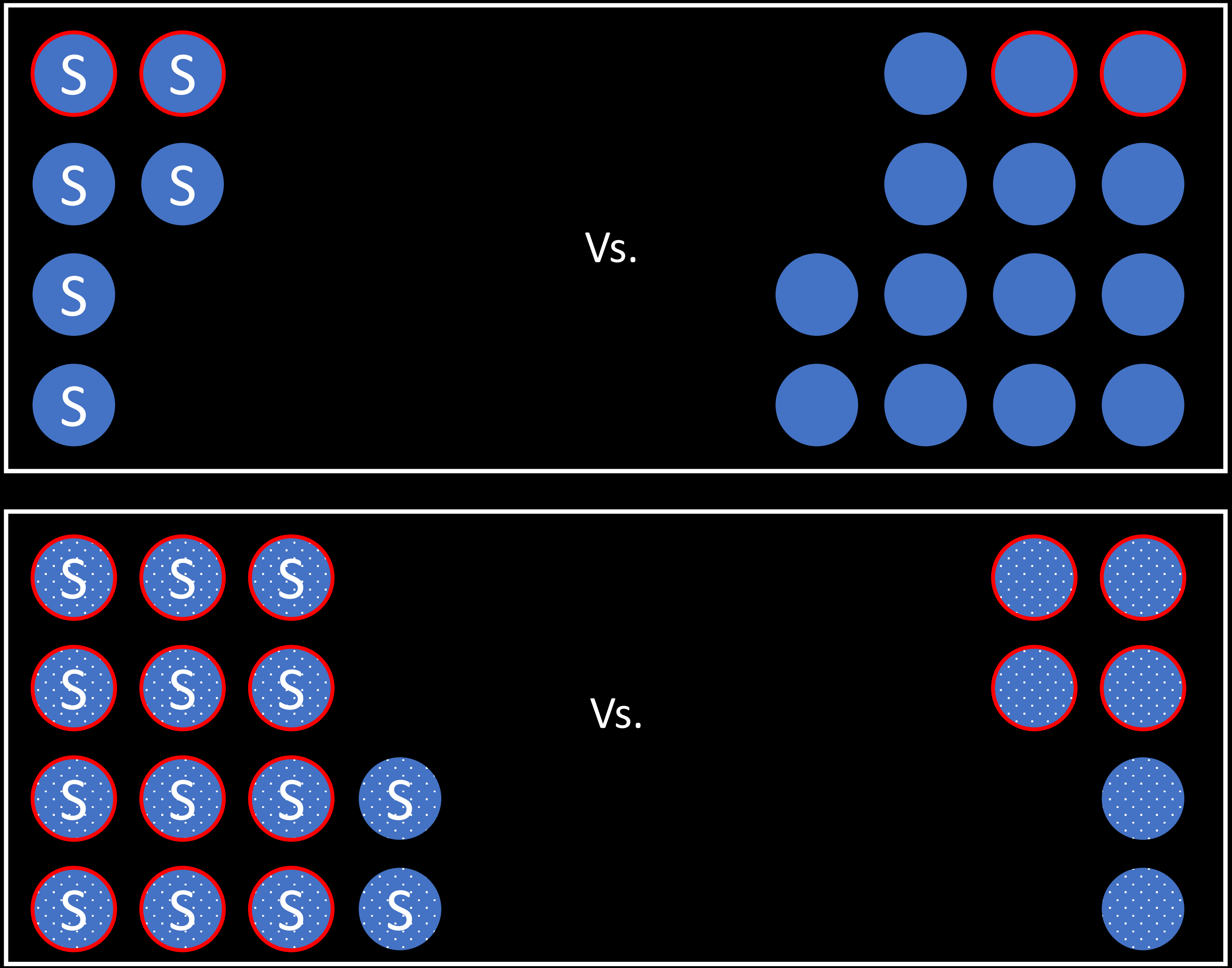


Vs.



Linear Regression

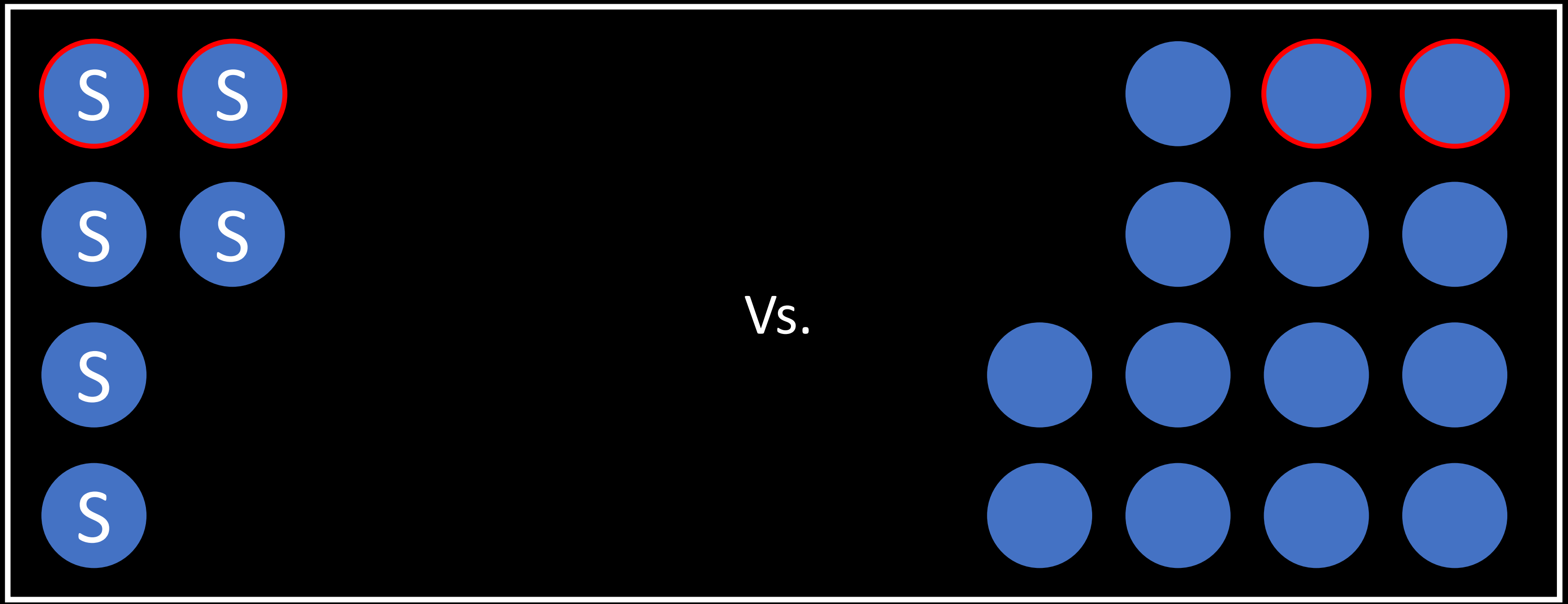
Linear Regression Stratifies the population



Keletso Makofane, June 2020

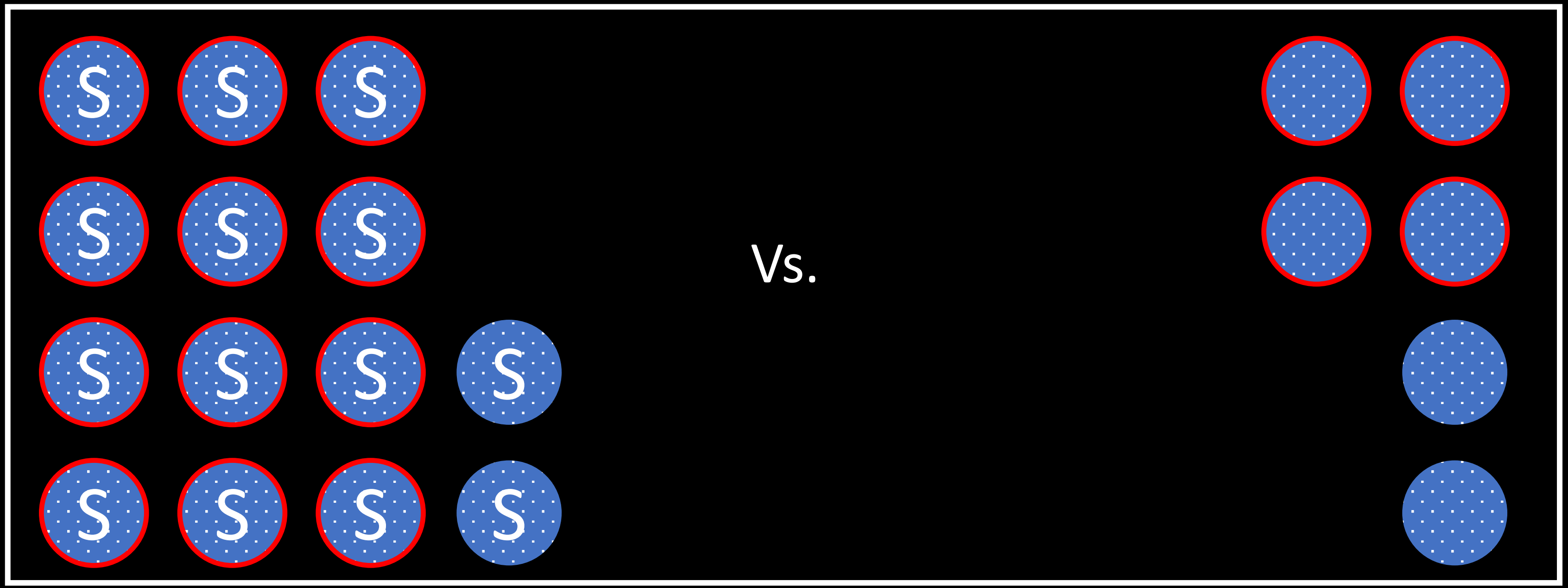
Effect of smoking among the unstressed

$$\frac{2}{6} - \frac{2}{14} = \frac{4}{21}$$



Effect of smoking among the stressed

$$\frac{12}{14} - \frac{4}{6} = \frac{4}{21}$$

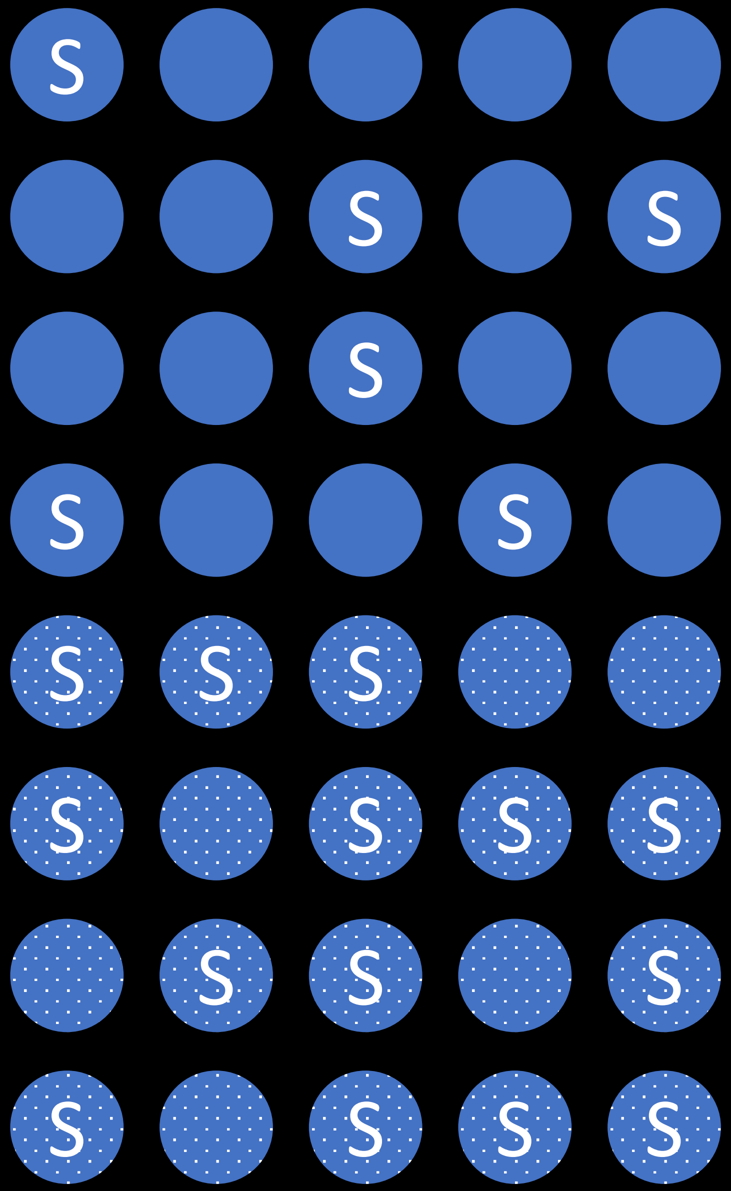


Keletso Makofane, June 2020

Regression is a model for the stratum-specific mean of the outcome

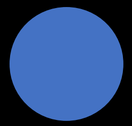
$$\hat{E}[Y|\, Smoking, Stress] = \frac{2}{14} + \frac{4}{21} Smoking + \frac{22}{42} Stress$$

Weighting for Single Treatment



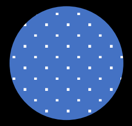
Imagine an RCT that assigns Smoking as follows:

$$P[Smoker = 1 \mid Stress = 0] = \frac{6}{20}$$

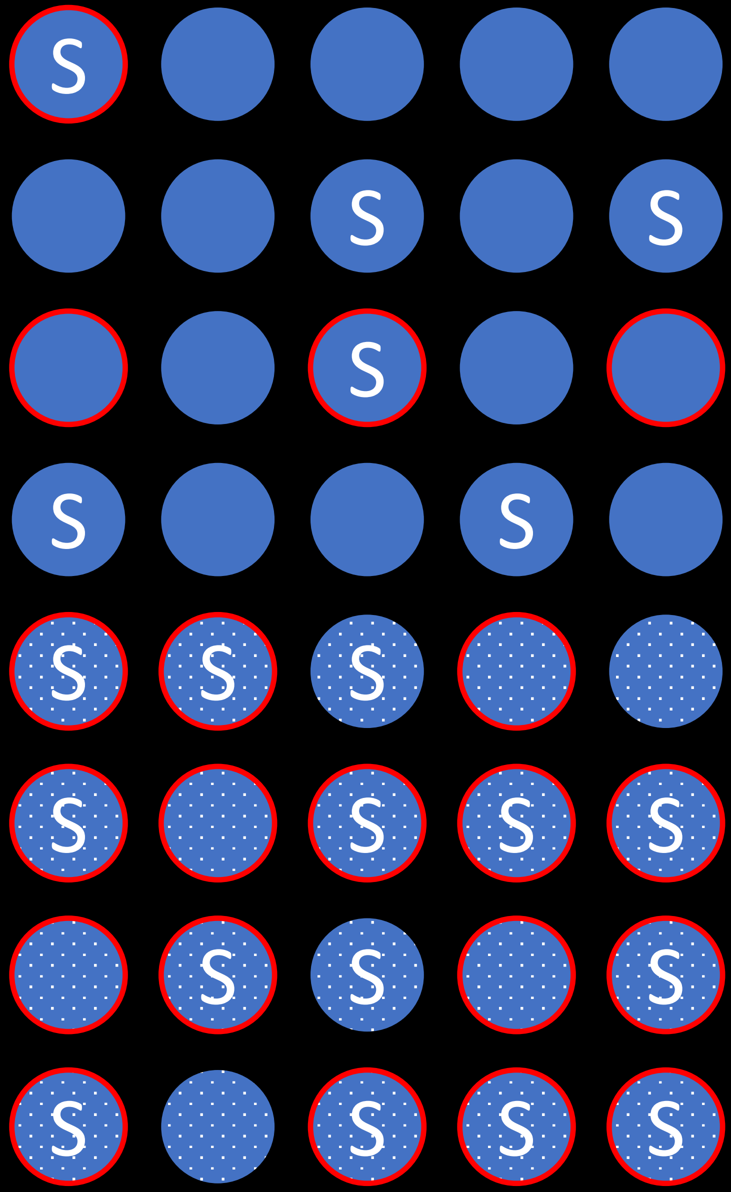


$$P[Smoker = 0 \mid Stress = 0] = \frac{14}{20}$$

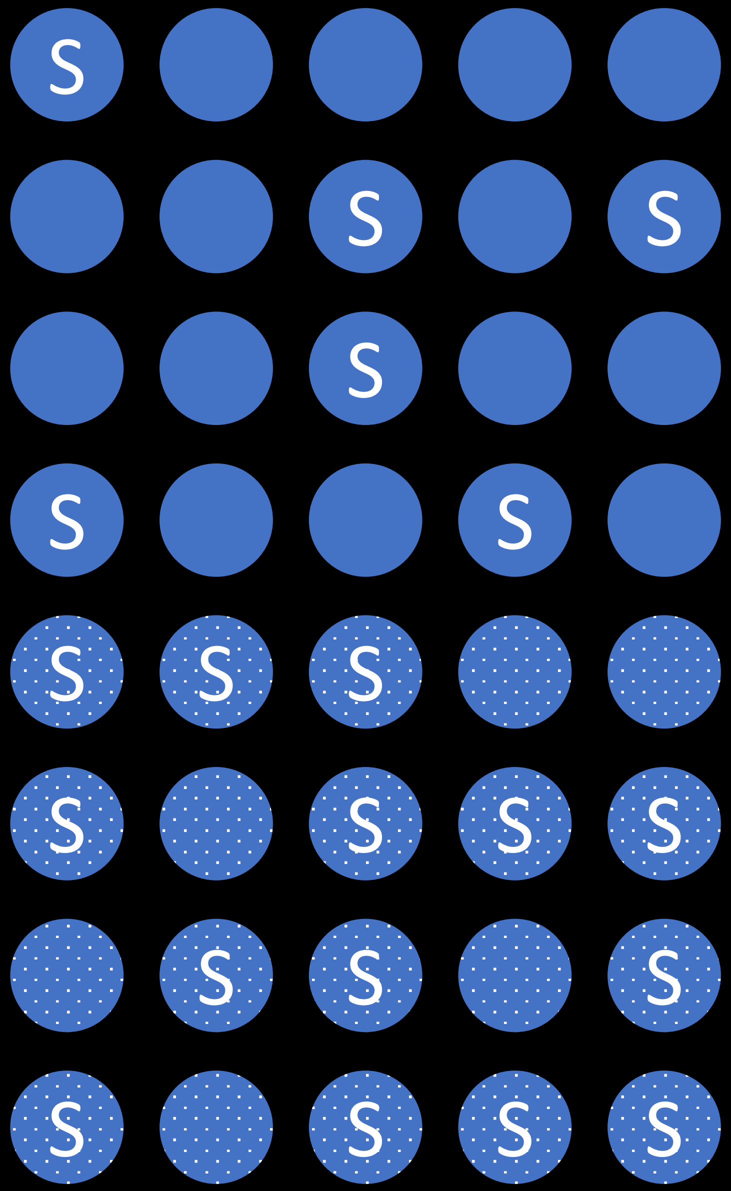
$$P[Smoker = 1 \mid Stress = 1] = \frac{14}{20}$$



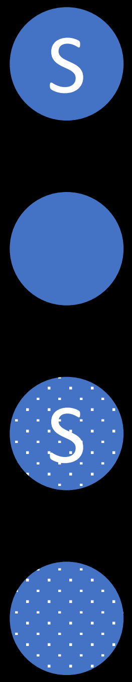
$$P[Smoker = 0 \mid Stress = 1] = \frac{6}{20}$$



The RCT follows the cohort and assesses lung cancer



We decide to undo the randomization scheme using weights

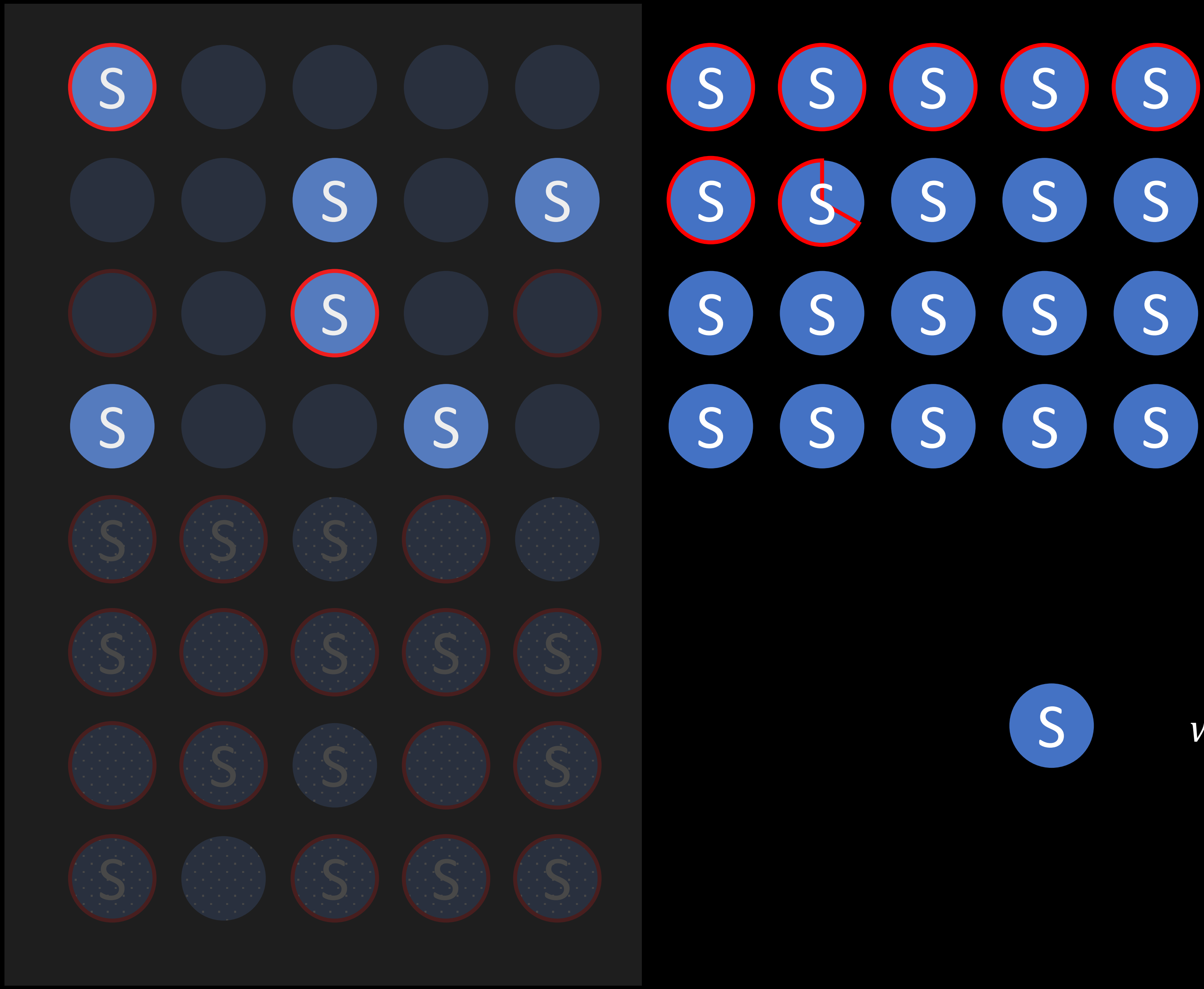


$$w = \frac{1}{P[Smoker = 1 \mid Stress = 0]} = \frac{20}{6}$$

$$w = \frac{1}{P[Smoker = 0 \mid Stress = 0]} = \frac{20}{14}$$

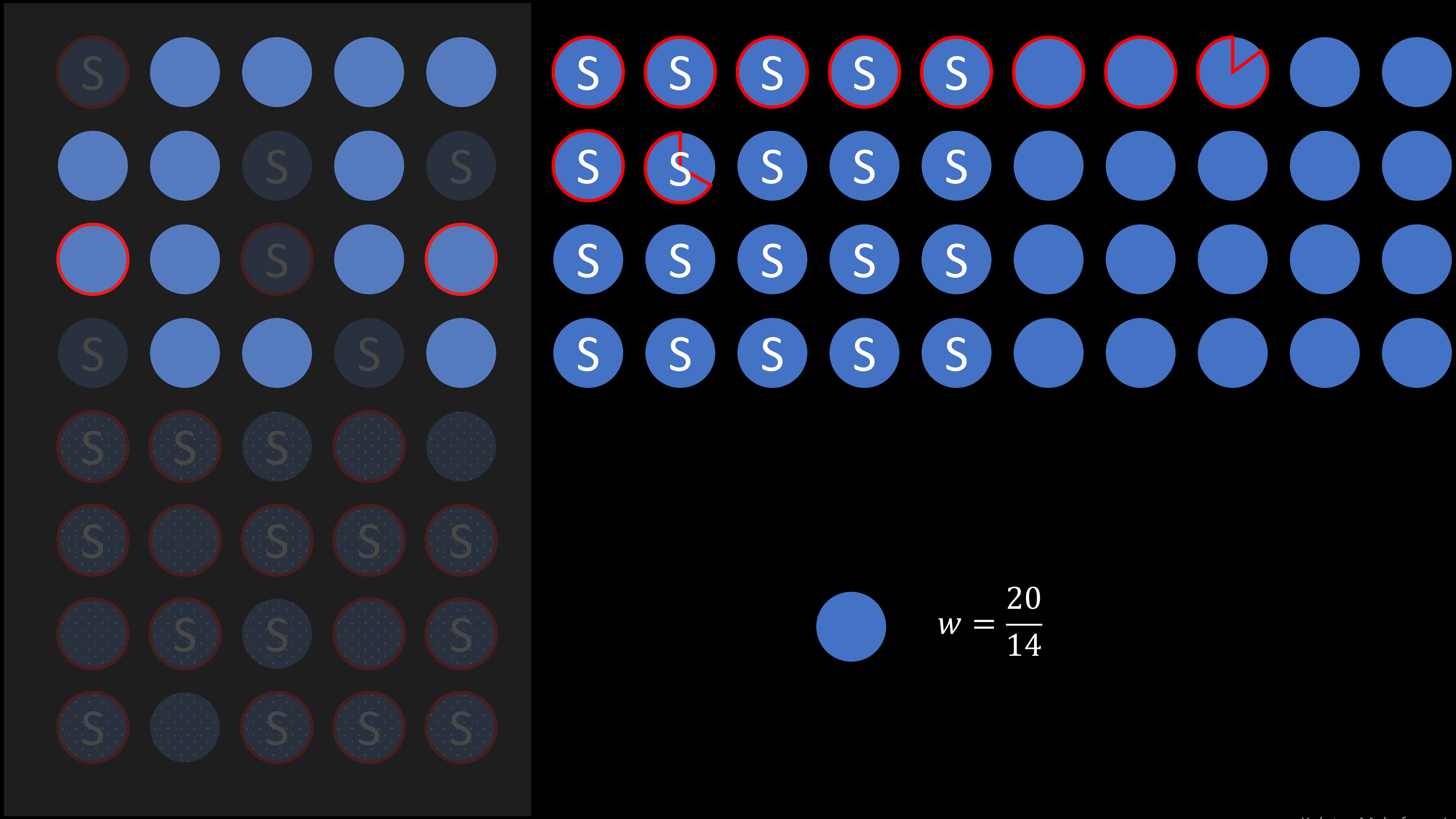
$$w = \frac{1}{P[Smoker = 1 \mid Stress = 1]} = \frac{20}{14}$$

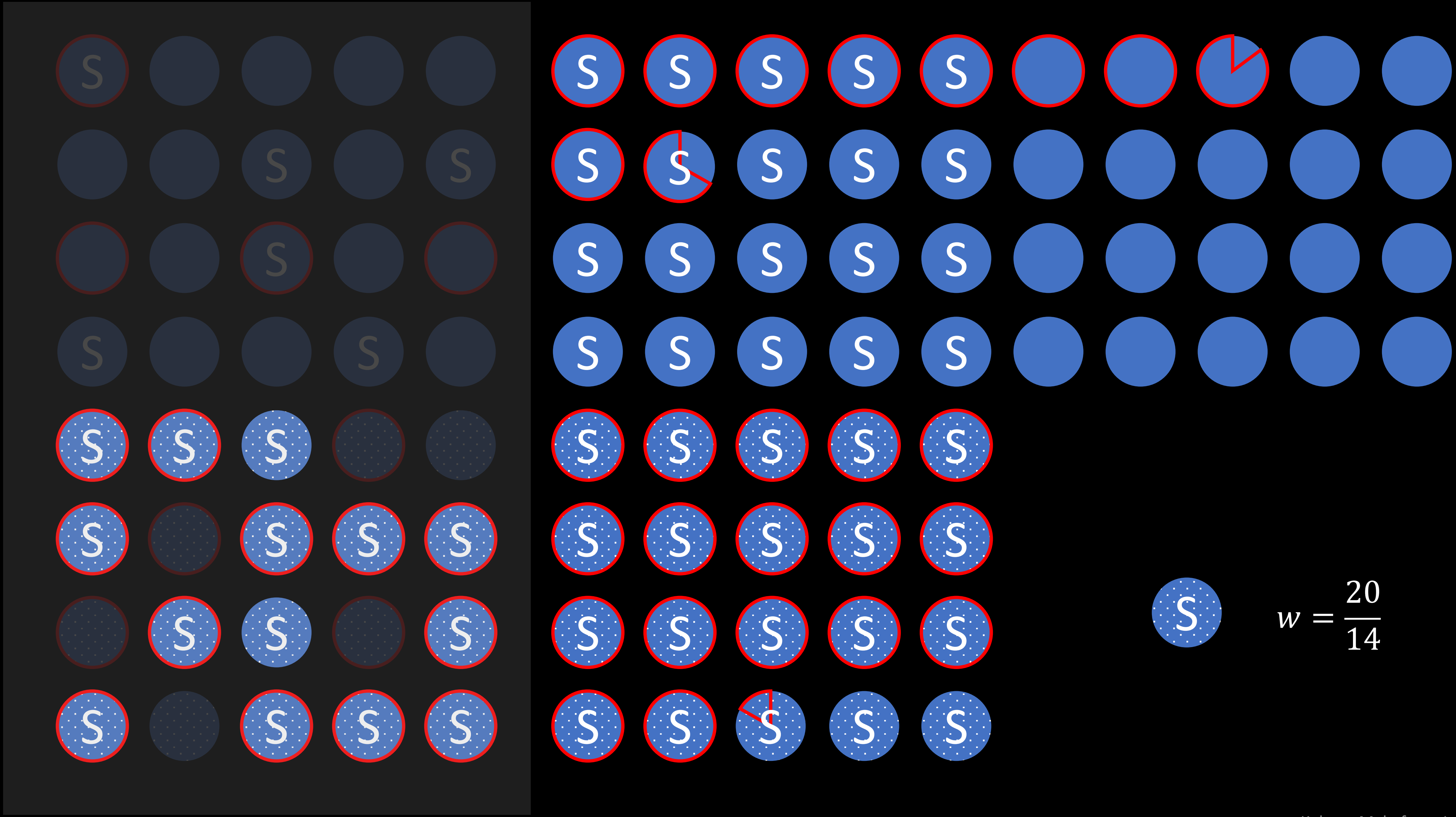
$$w = \frac{1}{P[Smoker = 0 \mid Stress = 1]} = \frac{20}{6}$$



$$w = \frac{20}{6}$$

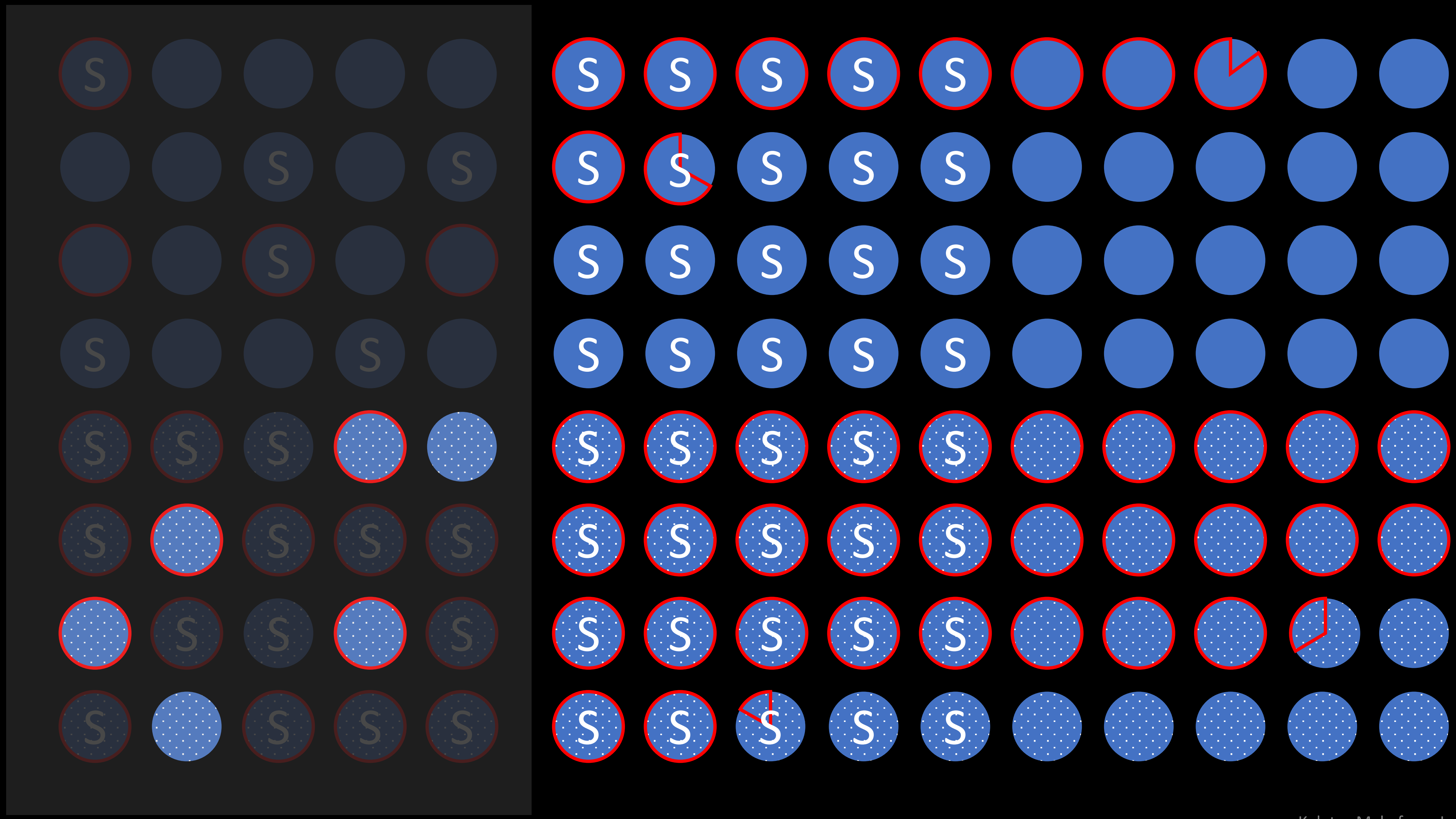
Keletso Makofane, June 2020


S
$w = \frac{20}{14}$



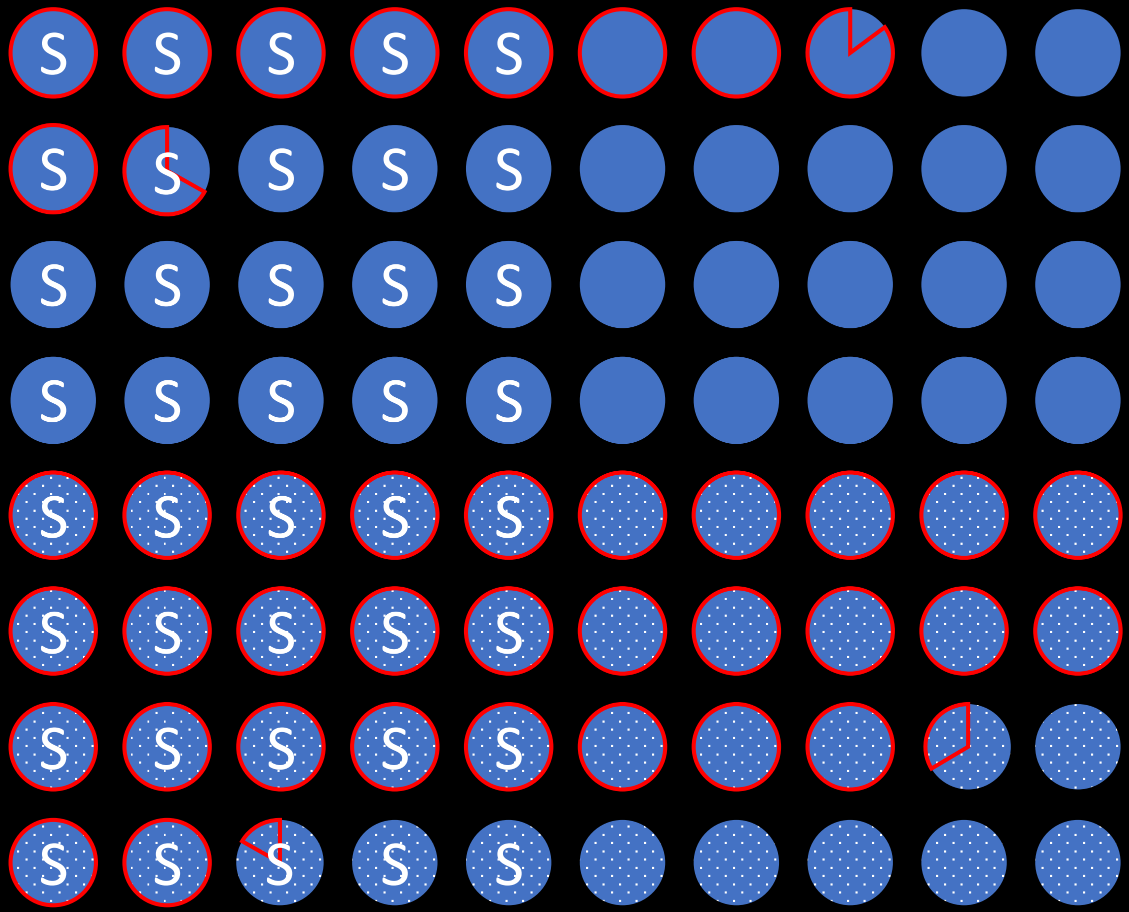
$$w = \frac{20}{14}$$

Keletso Makofane, June 2020

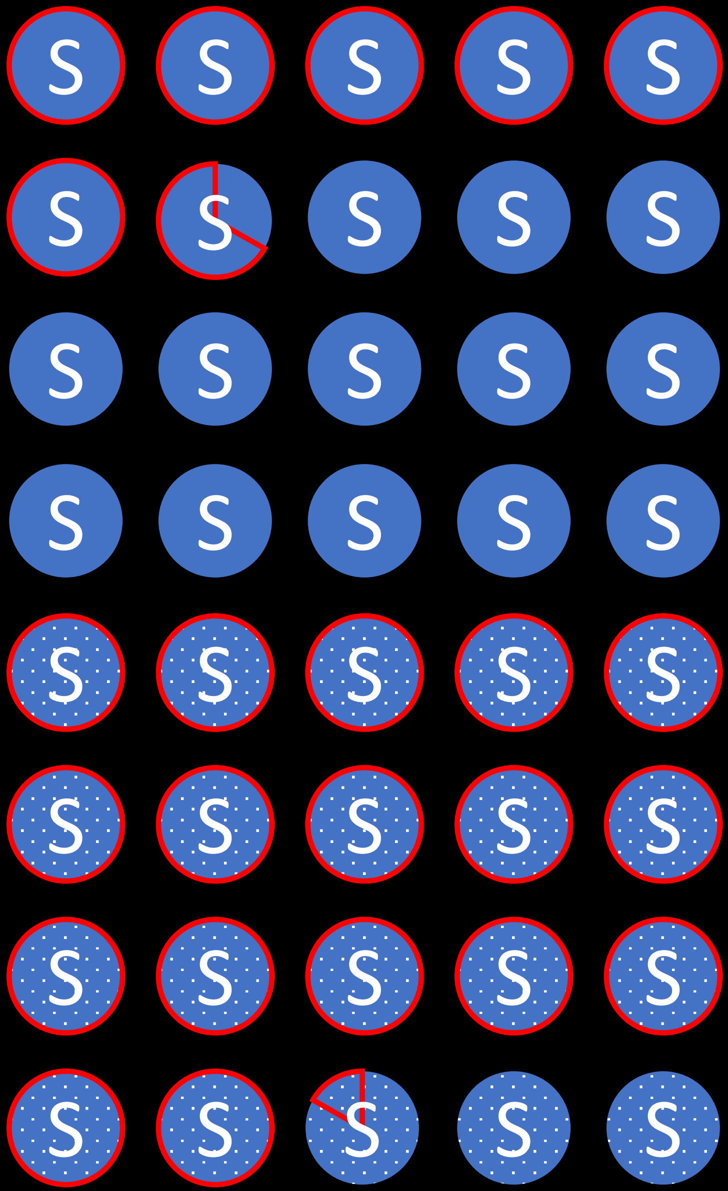


Keletso Makofane, June 2020

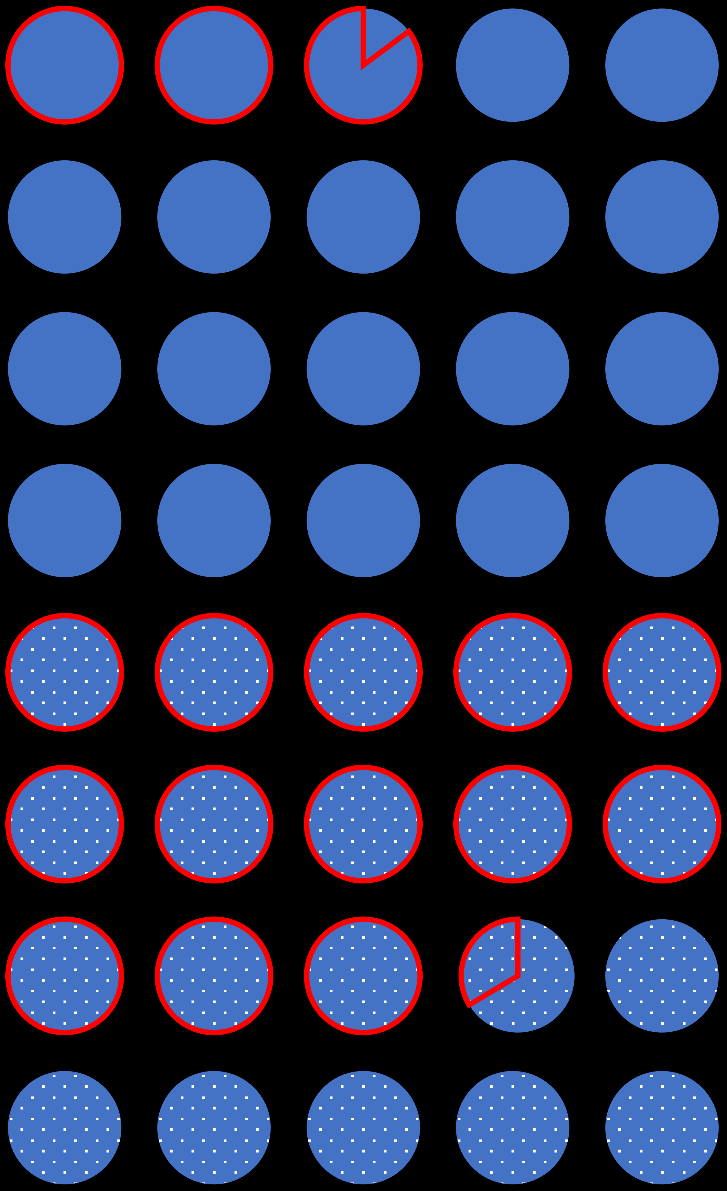
In our pseudo population, the distribution of stress is the same among smokers as it is among non-smokers



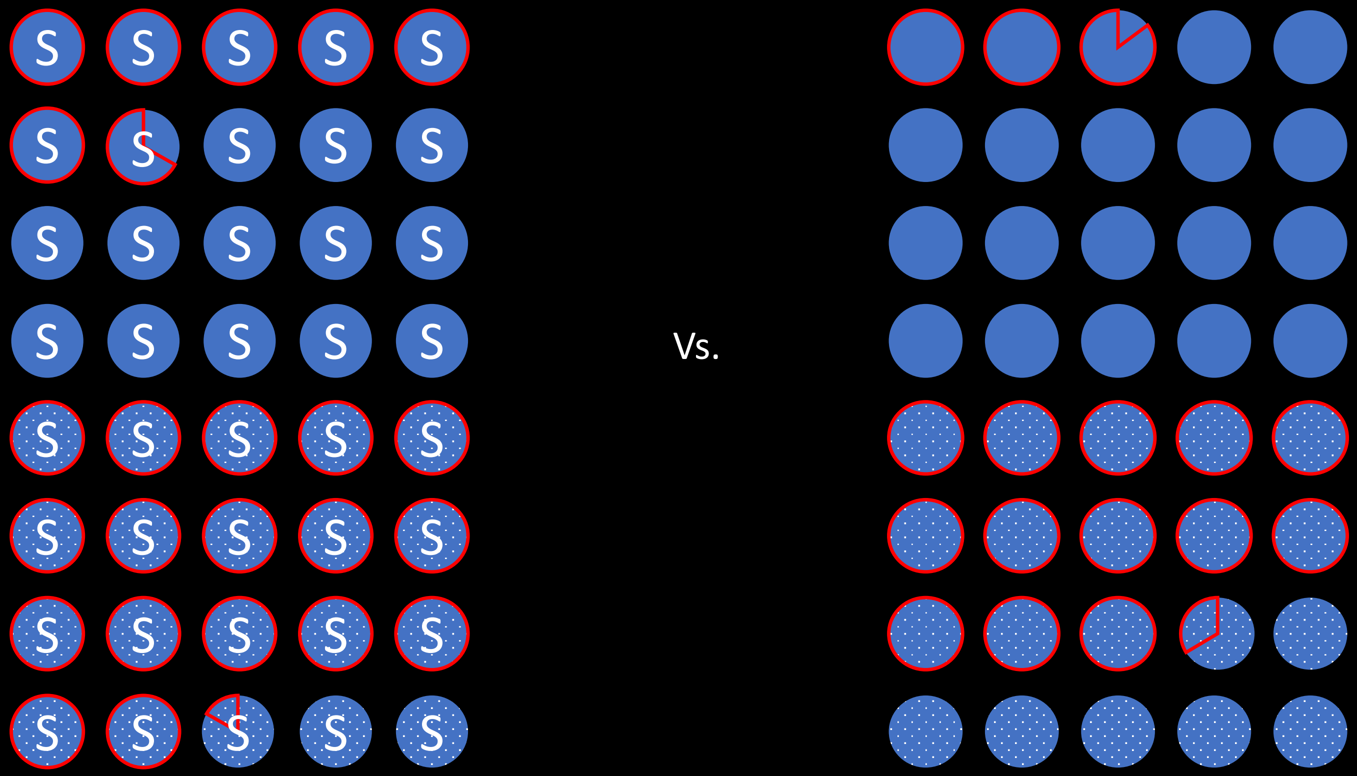
Keletso Makofane, June 2020



Vs.



Keletso Makofane, June 2020



Effect of smoking among the population:

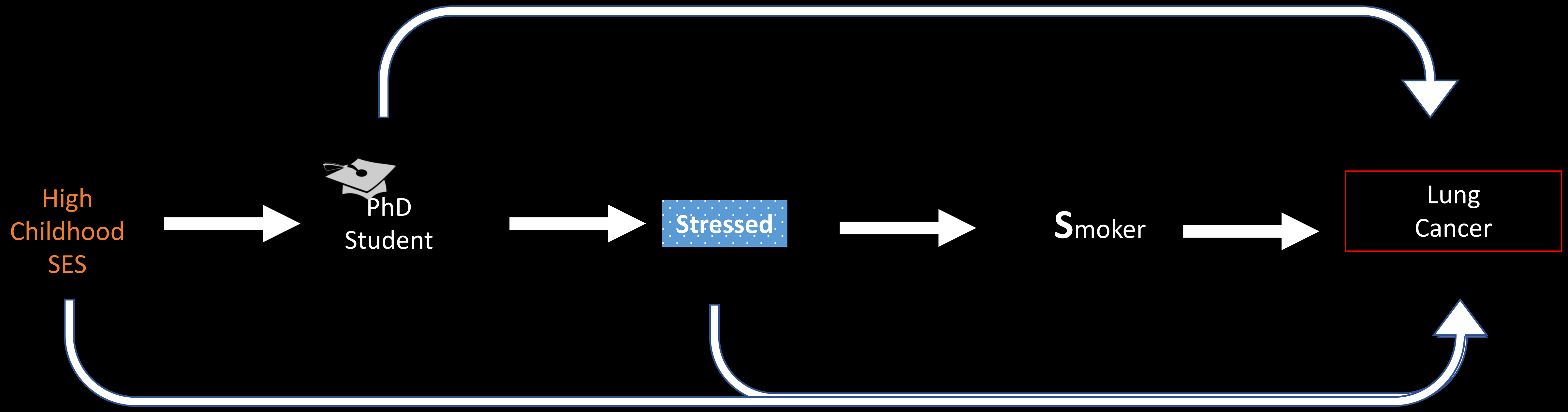
$$\frac{23.8095}{40} - \frac{16.1904}{40} = \frac{4}{21}$$

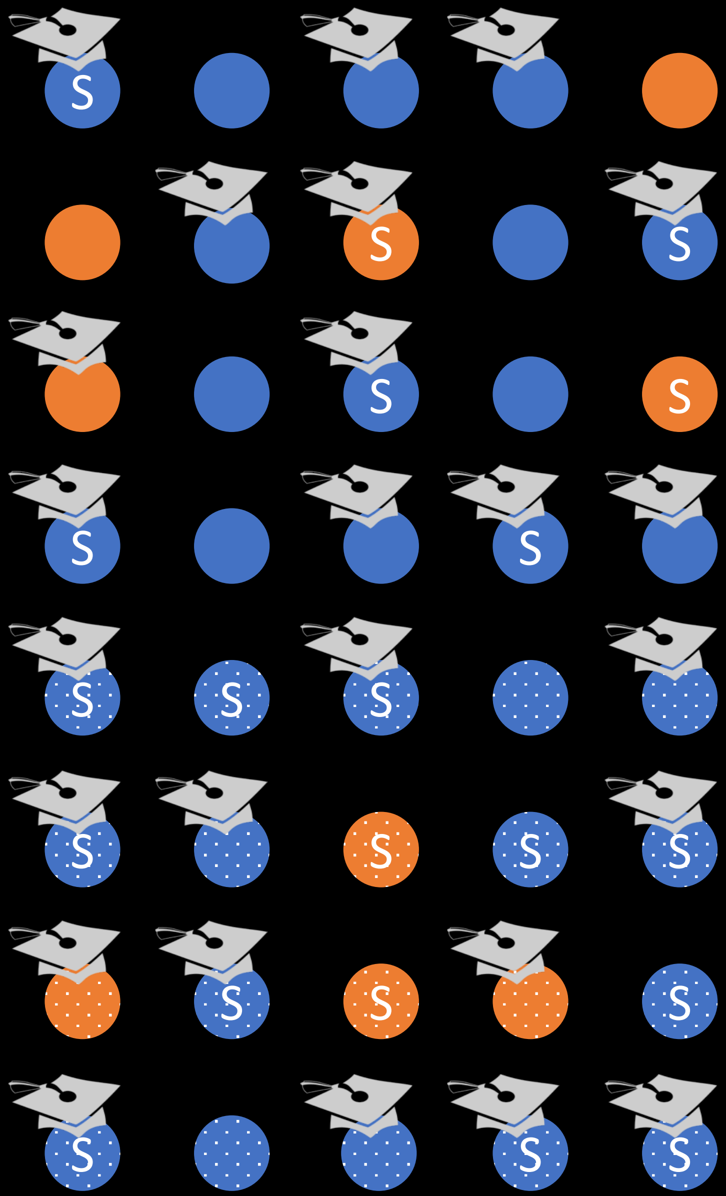
MSM is a model for the counterfactual mean of the outcome

$$\hat{E}[Y_{smoking}] = \frac{16.1904}{40} + \frac{4}{21}Smoking$$

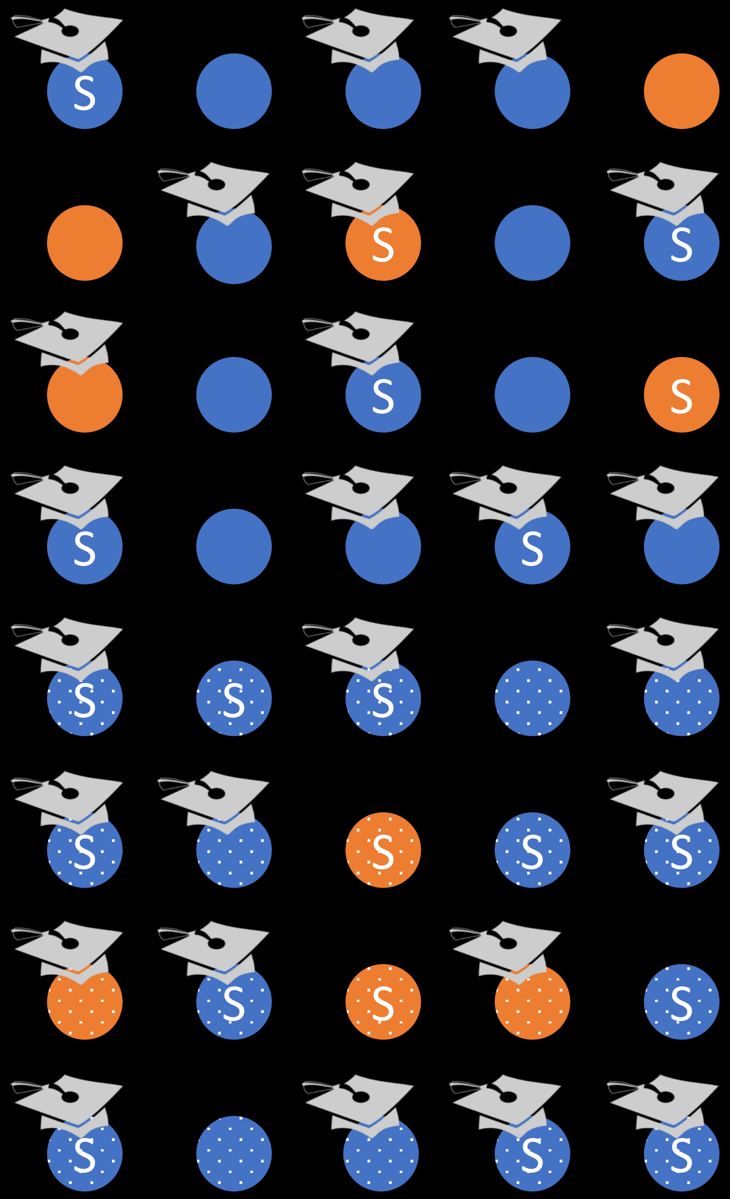
Weighting for Treatment at Multiple Times

We are interested in the joint effect of smoking and PhD studies on lung cancer

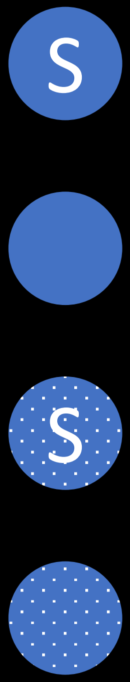




Keletso Makofane, June 2020



We imagine that Smoking was assigned as follows:

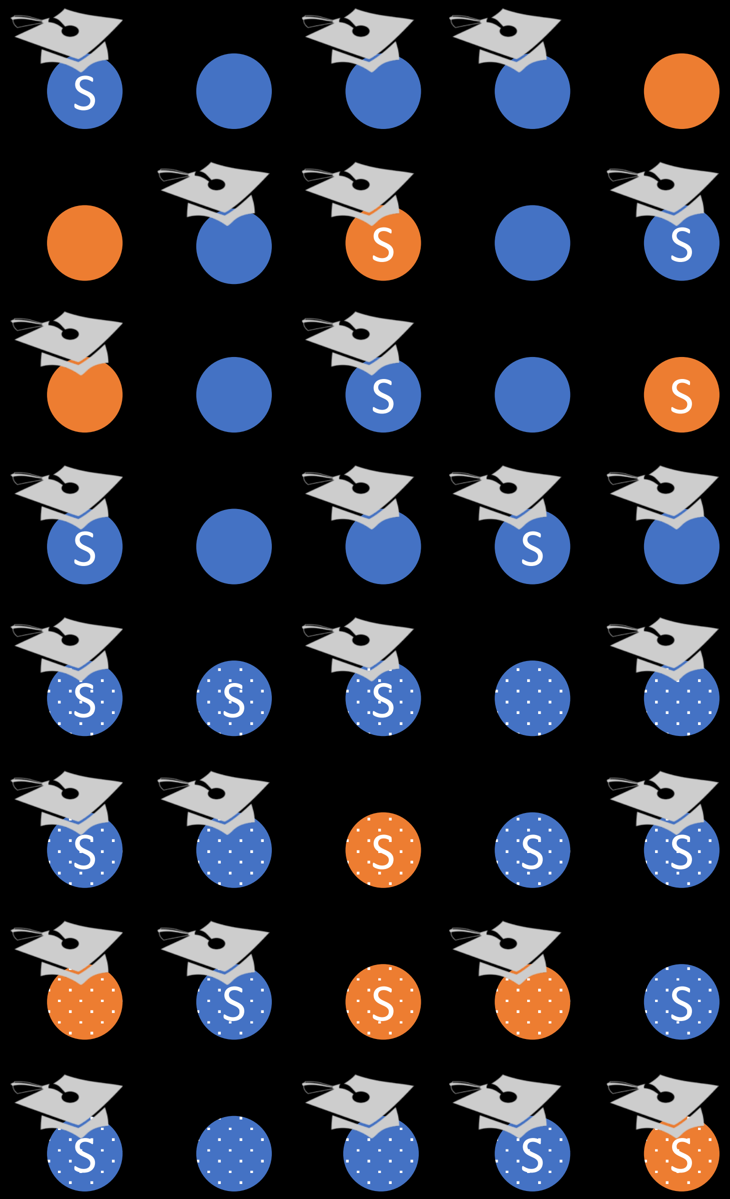


$$P[Smoker = 1 \mid Stress = 0, PhD, cSES]$$

$$P[Smoker = 0 \mid Stress = 0, PhD, cSES]$$

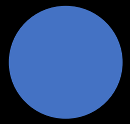
$$P[Smoker = 1 \mid Stress = 1, PhD, cSES]$$

$$P[Smoker = 0 \mid Stress = 1, PhD, cSES]$$



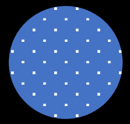
We undo this portion of the sequential randomization scheme using weights

$$w = \frac{1}{P[Smoker = 1 \mid Stress = 0, PhD, cSES]}$$

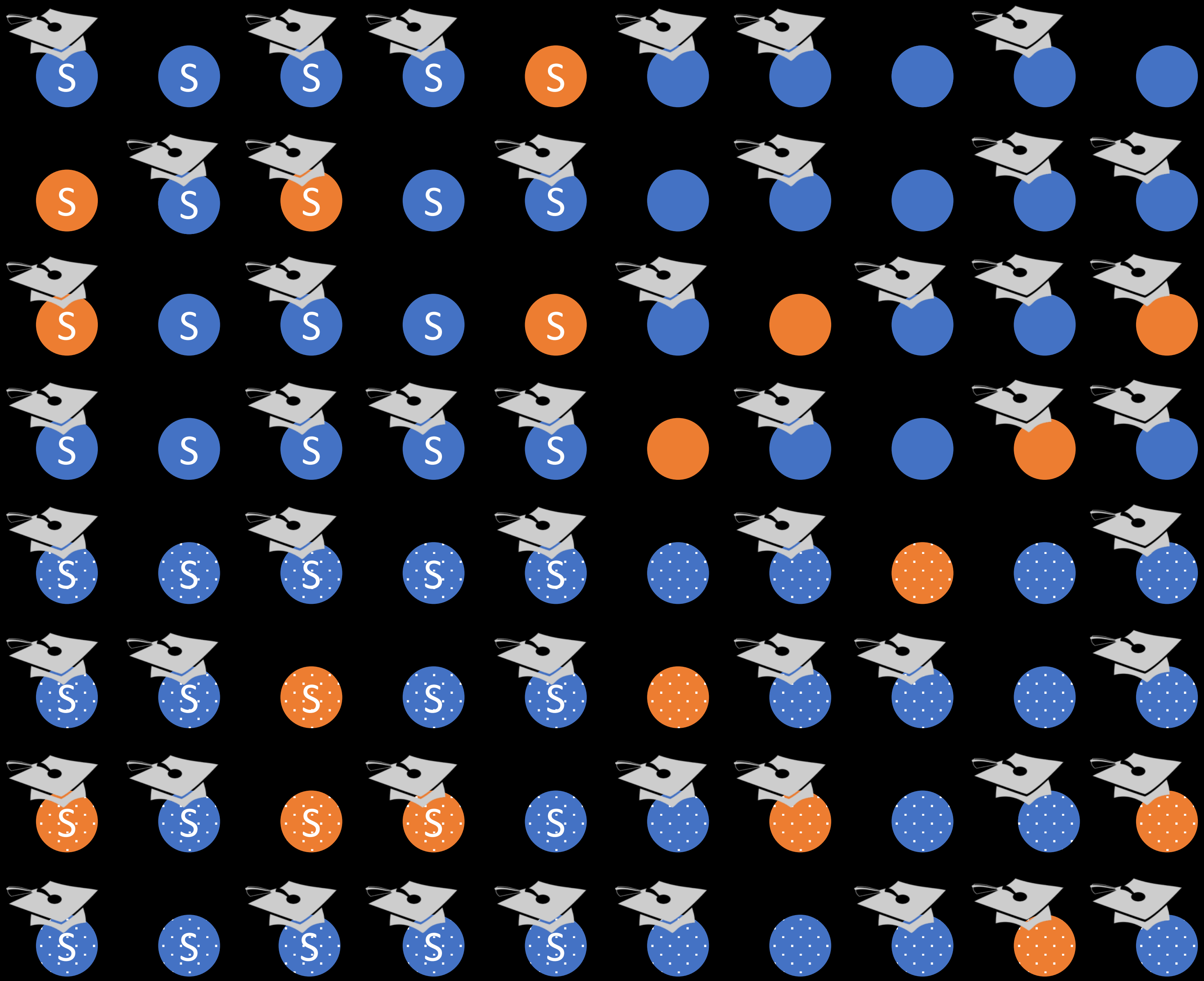


$$w = \frac{1}{P[Smoker = 0 \mid Stress = 0, PhD, cSES]}$$

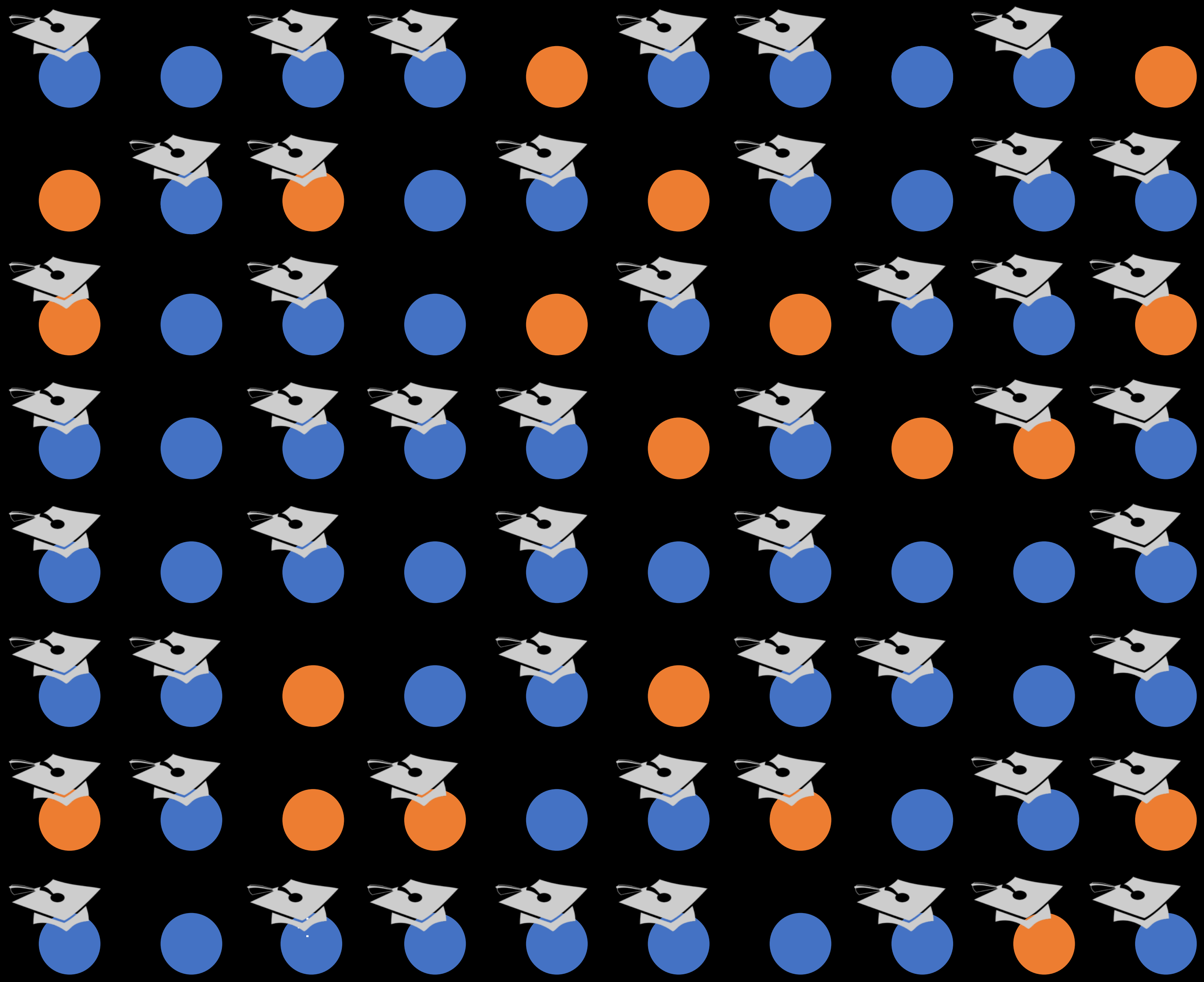
$$w = \frac{1}{P[Smoker = 1 \mid Stress = 1, PhD, cSES]}$$



$$w = \frac{1}{P[Smoker = 0 \mid Stress = 1, PhD, cSES]}$$



Keletso Makofane, June 2020



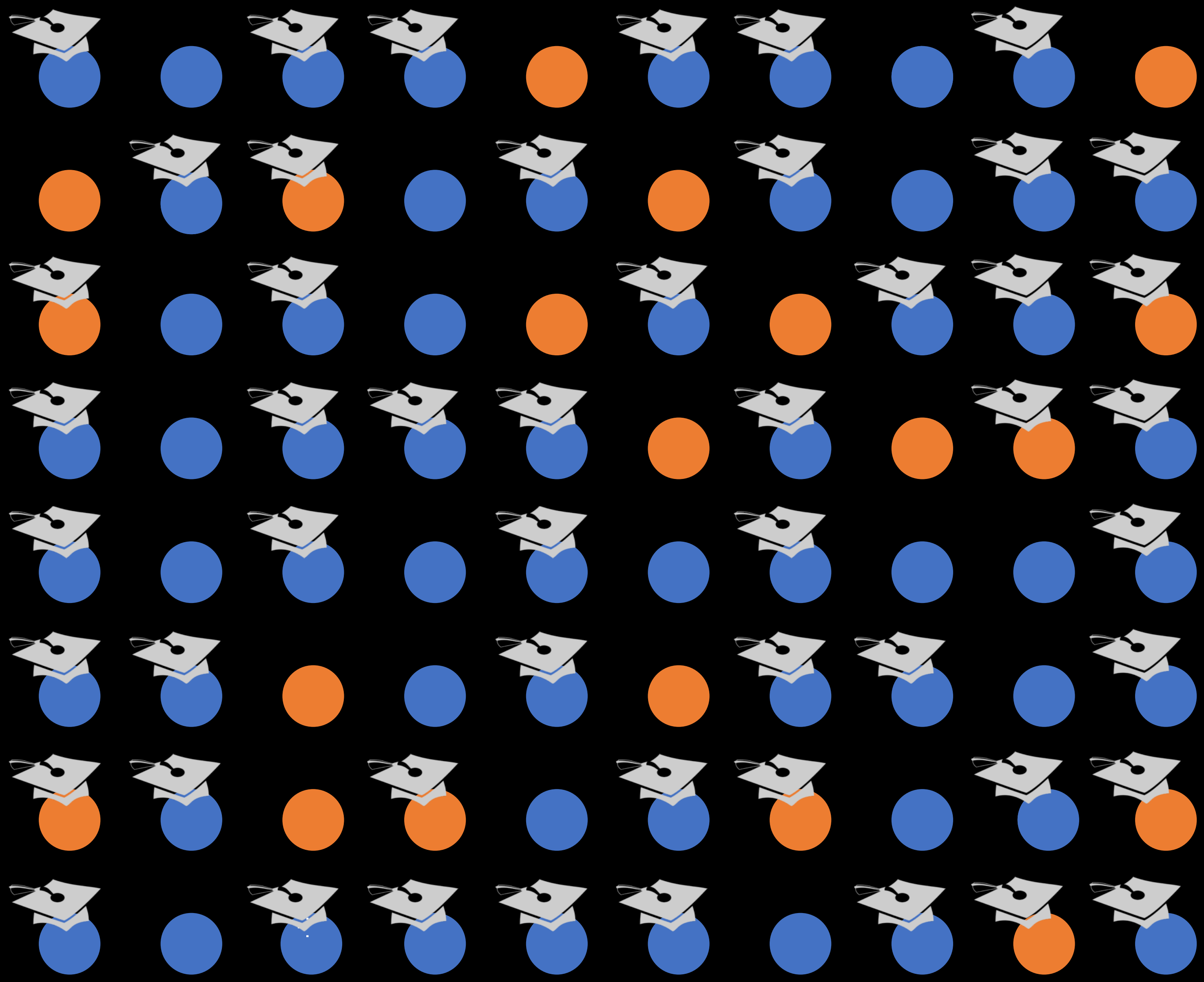
We imagine that PhD Status was assigned as follows:

$P[PhD = 1 \mid cSES = 0]$

$P[PhD = 0 \mid cSES = 0]$

$P[PhD = 1 \mid cSES = 1]$

$P[PhD = 0 \mid cSES = 1]$



We undo this portion of the sequential randomization scheme using weights

$$w = \frac{1}{P[PhD = 1 \mid cSES = 0]}$$

$$w = \frac{1}{P[PhD = 0 \mid cSES = 0]}$$

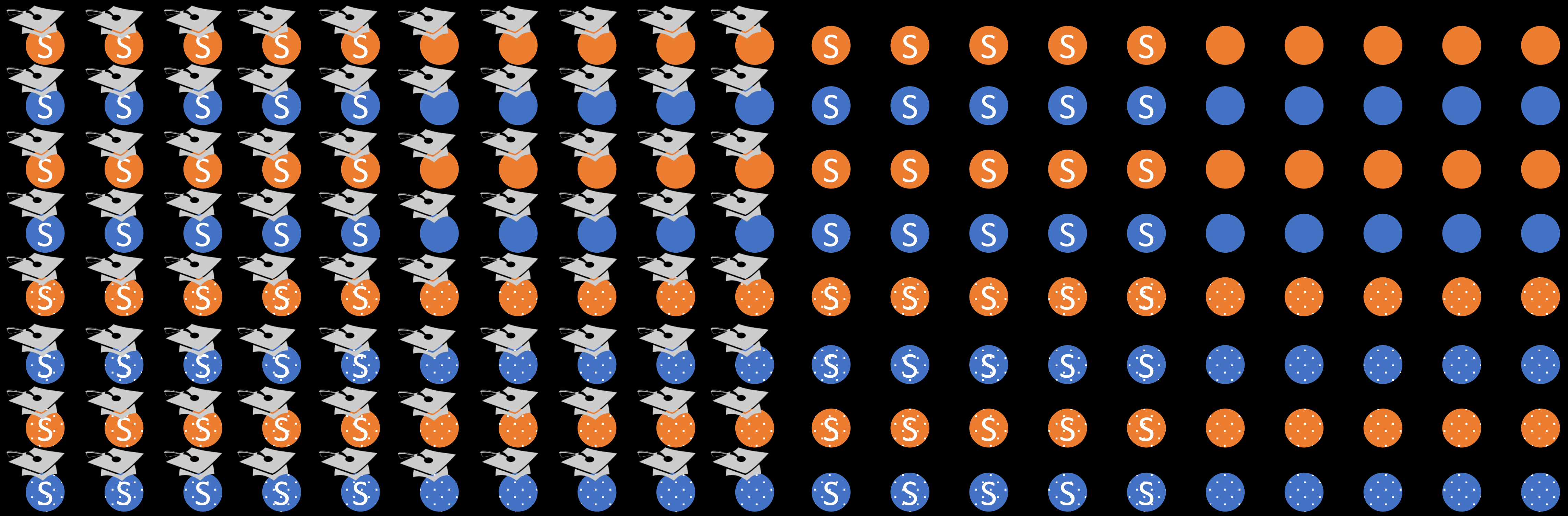
$$w = \frac{1}{P[PhD = 1 \mid cSES = 1]}$$

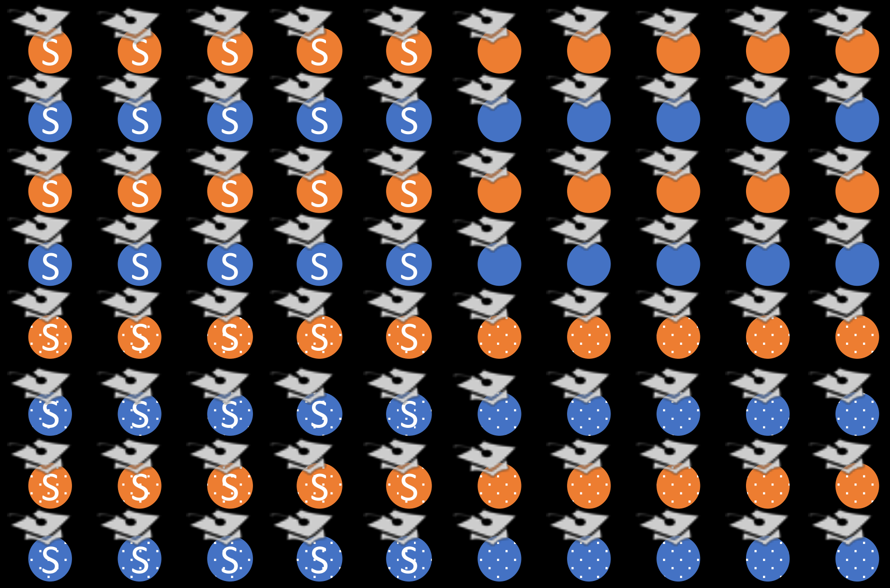
$$w = \frac{1}{P[PhD = 0 \mid cSES = 1]}$$

In our pseudo population:

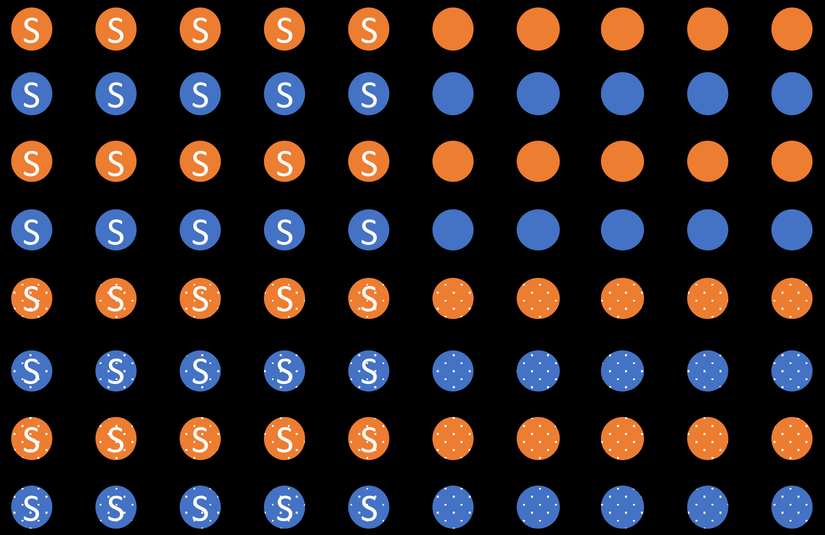
a) the distribution of stress and cSES is the same among smokers as it is among non-smokers

b) the distribution of stress and cSES is the same among PhD students as it is among non-PhD students



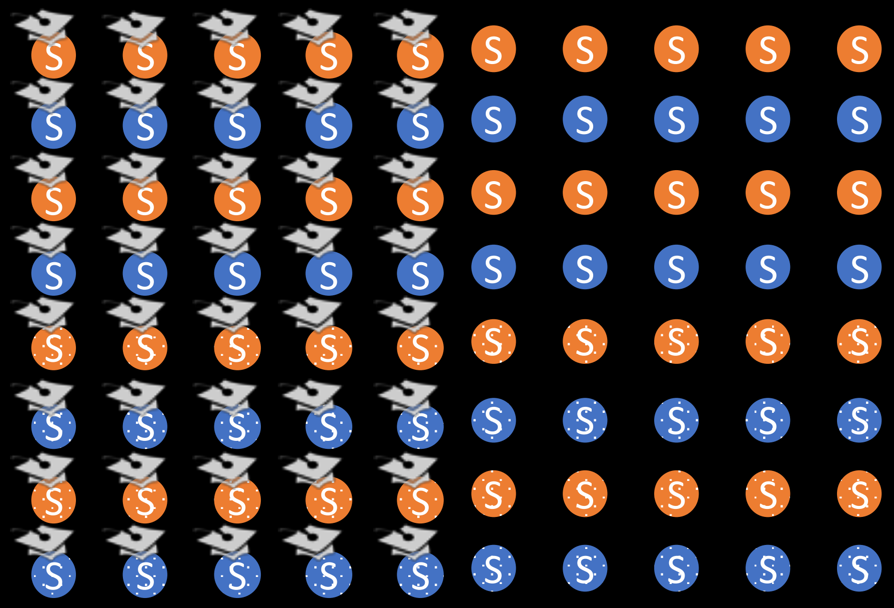


Vs.

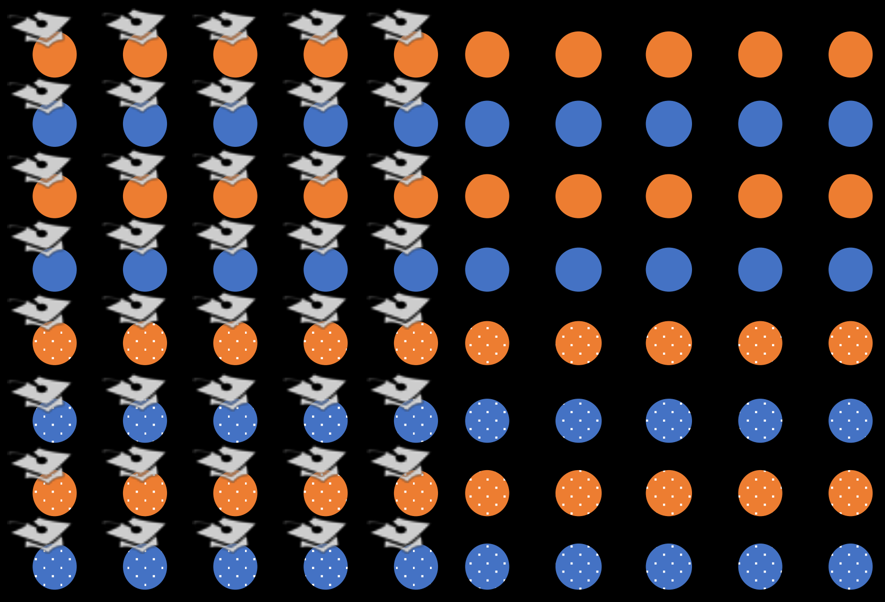


Effect of PhD Studies among the population:

$$\text{Cumulative Incidence of Lung Cancer}_{\text{LHS}} - \text{Cumulative Incidence of Lung Cancer}_{\text{RHS}} = \hat{\beta}_{PhD}$$



Vs.



Effect of Smoking among the population:

$$\text{Cumulative Incidence of Lung Cancer}_{\text{LHS}} - \text{Cumulative Incidence of Lung Cancer}_{\text{RHS}} = \hat{\beta}_{Smoking}$$

MSM is a model for the counterfactual mean of the outcome

$$\hat{E}\left[Y_{smoking,PhD}\right] = \hat{\beta}_0 + \hat{\beta}_{smoking} Smoking + \hat{\beta}_{PhD} PhDStatus$$