Figure 1: Data Generation Mechanism and Parameter Values

1. Simulate $n$=1,000,000 individuals.

<u>Gender and Race</u>

$$\pi_G = P[G_i = 1] = 0.01, \pi_B = P[B_i = 1] = 0.134$$

<u>True Outcome Model</u>

$$\pi_{Y,i} = expit[\beta_0 + \beta_1 G_i + \beta_2 B_i + \beta_3 G_i * B_i]$$
$$Y_i \sim Bernouilli\ (\pi_{Y,i})$$
$$OR_1 = \exp(\beta_1)\ , OR_2 = \exp(\beta_1 + \beta_2 + \beta_3), OR_3 = \exp\ (\beta_2 + \beta_3)$$

<u>Misclassification Model</u>

$$\pi_{M,i} = G * expit[\alpha_0 + \alpha_1 B_i + \alpha_2 Y_i]$$
$$M_i \sim Bernouilli\ (\pi_{M,i})$$
$$G_i^* = M_i(1 - G_i) + (1 - M_i)G_i$$

3. Repeat $N$=1,000 times to obtain the simulated distribution of the observed ORs.

<u>Observed Outcome Model</u>

$$logit[P(Y = 1)] = \gamma_0 + \gamma_1 G^* + \gamma_2 B + \gamma_3 G^* * B$$
$$\widehat{OR}_1 = \exp(\gamma_1)\ , \widehat{OR}_2 = \exp(\gamma_1 + \gamma_2 + \gamma_3), \widehat{OR}_3 = \exp\ (\gamma_2 + \gamma_3)$$

| | Parameter Values | Probability of Misclassification/Odds Ratios |
|---|---|---|
| Mild MC | $\alpha_0 = -2.18, \alpha_1 = 1.5, \alpha_2 = 0.45$ | TW=0.10, TWY=0.15, TB=0.34, TBY=0.44 |
| Moderate MC | $\alpha_0 = -0.85, \alpha_1 = 1.65, \alpha_2 = 0.45$ | TW=0.30, TWY=0.40, TB=0.69, TBY=0.78 |
| Severe MC | $\alpha_0 = 0, \alpha_1 = 1.8, \alpha_2 = 0.45$ | TW=0.50, TWY=0.61, TB=0.87, TBY=0.91 |
| Mild Effect | $\beta_0 = -2.5, \beta_1 = 0.41, \beta_2 = 0.27, \beta_3 = 0.24$ | $OR_1$=1.5, $OR_2$=2.5, $OR_3$= 1.7 |
| Moderate Effect | $\beta_0 = -2.5,\ \beta_1 = 0.7, \beta_2 = 0.24, \beta_3 = 0.24$ | $OR_1$=2.0, $OR_2$=3.8, $OR_3$=1.9 |
| Strong Effect | $\beta_0 = -2.5,\ \beta_1 = 1.1, \beta_2 = 0.26, \beta_3 = 0.15$ | $OR_1$=3.0 $OR_2$=6.8, $OR_3$=2.3 |