

## Appendix

### Inverse Probability Weighted Marginal Structural Models

Predicted probabilities used in exposure and mediator weights were created with logistic regression models.

$$w_i^{X*} w_i^{M} = \frac{P(X=x_i)}{P(X=x_i|C_{XY}=c_i)} * \frac{P(M=m_i|X=x_i)}{P(M=m_i|X=x_i, C_{XY}=c_i, C_{MY}=l_i)} \quad (1)$$

In formula (1), the denominator used to calculate the weight for the exposure, history of vaginal douching (< 2 times in past month; ≥2 times in past month), was created by a logistic regression model with vaginal douching predicted by the exposure-outcome confounders, age at baseline (<25 years; ≥25 years) and race/ethnicity (African-American; Caucasian/Hispanic/Other).

The weights for the mediators, endometrial infection with one or more BV associated organisms (Positive; Negative) and endometrial infection with all BV associated organisms (Positive; Negative), were each calculated with logistic regression models providing the predicted probability of the mediator adjusting for vaginal douching in the numerator, and the predicted probability of the mediator adjusting for vaginal douching and race/ethnicity, the mediator-outcome confounder, in the denominator.

$$\log [E(Y_{XM})] = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 X * M \quad (2)$$

The inverse of the product of the weights was applied to log-binomial marginal structural models (formula (2)) estimating the controlled direct effect of a history of vaginal douching on endometritis (Positive; Negative).

#### Reference:

VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009;20(1):18-26.