Web-based supporting materials for

Indices of Non-Ignorable Selection Bias for Proportions

Estimated from Non-Probability Samples

by

Andridge, West, Little, Boonstra, and Alvarado-Leiton

Proof of Reduced Proxy Pattern-Mixture Formulation – Reducing Equation (5) to (7)

To show that (5) reduces to (7) under the stated assumptions, note first that since *V* is assumed to be uncorrelated with *X* for both selected and non-selected cases, $\beta_{xv\cdot v}^{(0)} = \beta_{xv\cdot v}^{(1)} = 0$. Secondly, since *X* is assumed to be the best predictor of *Y* for both selected and non-selected cases,

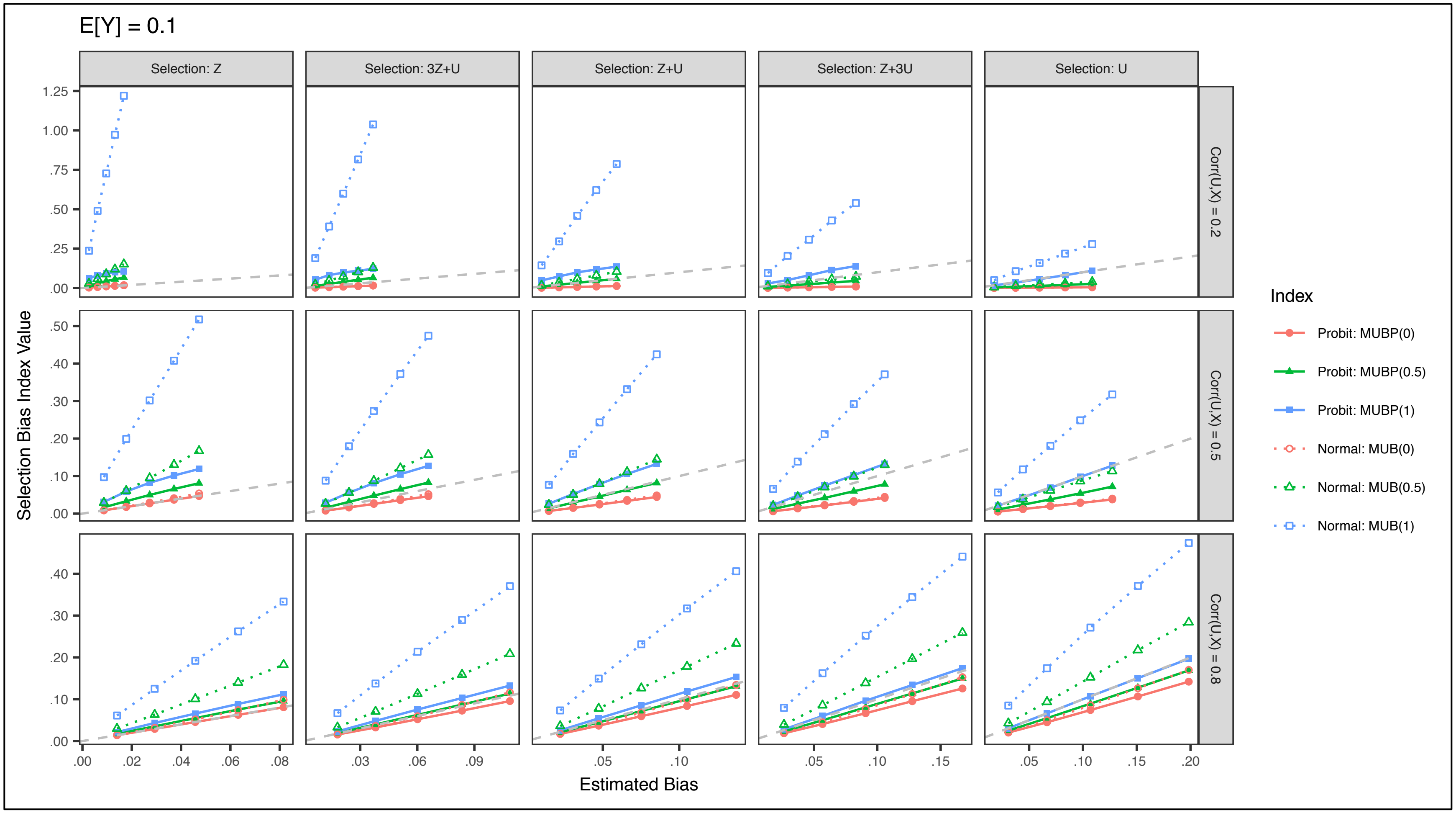
$$\begin{aligned} \beta_{uv\cdot vx}^{(j)} &= \beta_{uv\cdot v}^{(j)} - \frac{\sigma_{ux\cdot v}^{(j)}\beta_{xv\cdot v}^{(j)}}{\sigma_{xx\cdot v}^{(j)}} \\ &= \beta_{uv\cdot v}^{(j)} = 0 \end{aligned} \tag{A1}$$

for *j* = 0, 1. Thus the model in (5) reduces to

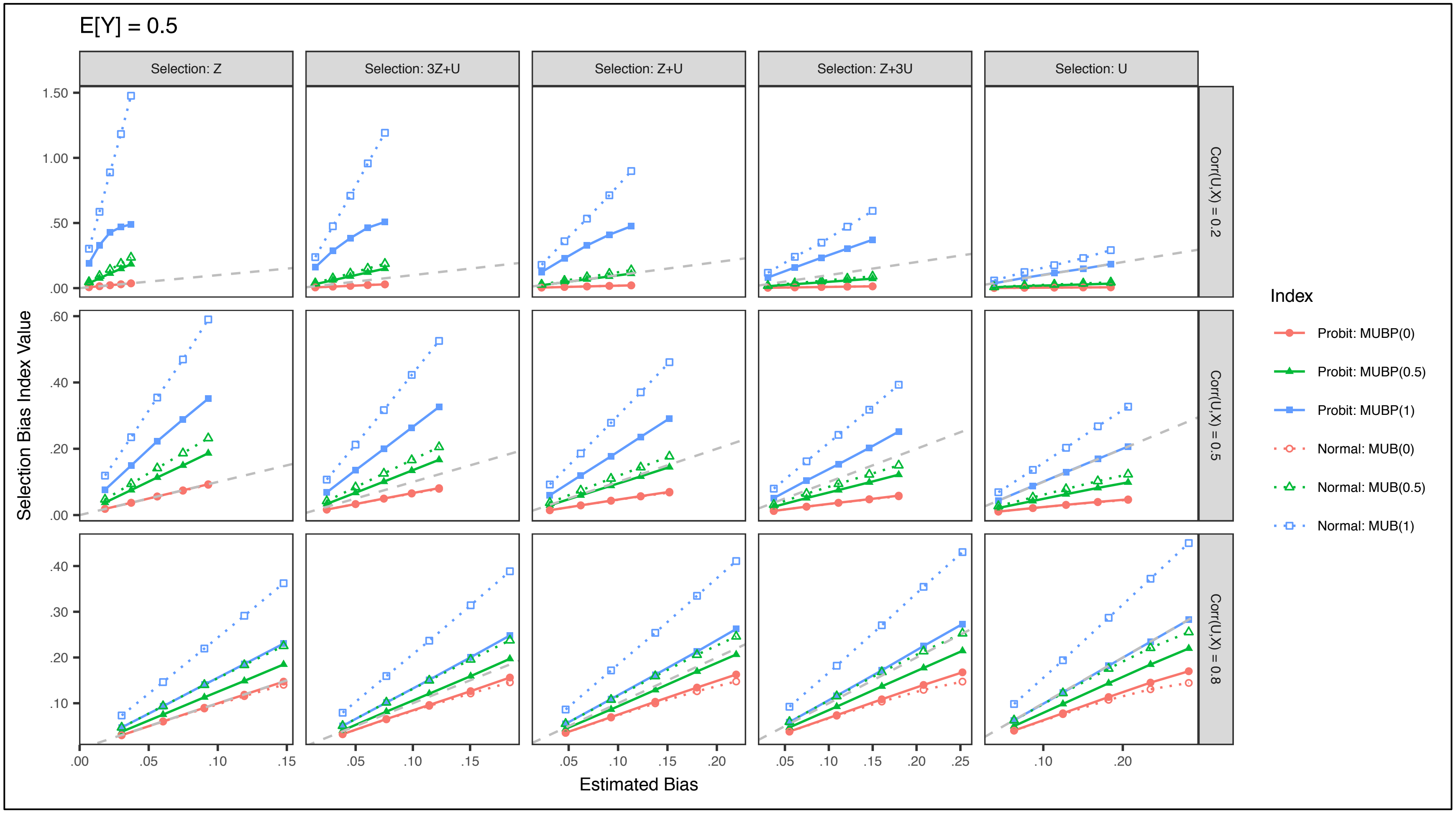
$$(U, X|V, S=j) \sim N_2\left(\begin{pmatrix}\beta_{u0\cdot v}^{(j)} \\ \beta_{x0\cdot v}^{(j)}\end{pmatrix}, \begin{pmatrix}\sigma_{uu\cdot v}^{(j)} & \sigma_{ux\cdot v}^{(j)} \\ \sigma_{ux\cdot v}^{(j)} & \sigma_{xx\cdot v}^{(j)}\end{pmatrix}\right), \tag{A2}$$

which is the pattern-mixture model in (7) with $\mu_u^{(j)} = \beta_{u0\cdot v}^{(j)}$, $\mu_x^{(j)} = \beta_{x0\cdot v}^{(j)}$, $\sigma_{uu}^{(j)} = \sigma_{uu\cdot v}^{(j)}$, $\sigma_{xx}^{(j)} = \sigma_{xx\cdot v}^{(j)}$, and $\rho_{ux}^{(j)} = \frac{\sigma_{ux}^{(j)}}{\sqrt{\sigma_{uu}^{(j)}\sigma_{xx}^{(j)}}} = \frac{\sigma_{ux\cdot v}^{(j)}}{\sqrt{\sigma_{uu\cdot v}^{(j)}\sigma_{xx\cdot v}^{(j)}}}$.

Supplemental Figure 1: MUBP(ϕ) from the probit model (solid lines/solid symbols) and MUB(ϕ) from the normal model (dotted lines/open symbols) versus the true estimated bias, shown for combinations of the biserial correlation $Corr(U, X) = \rho_{ux}$ (rows) and the selection mechanism (columns), for E[Y] = 0.1. Grey dashed line is equality (index = estimated bias). Results are medians across 1000 simulated data sets for each scenario.



Supplemental Figure 2: MUBP(ϕ) from the probit model (solid lines/solid symbols) and MUB(ϕ) from the normal model (dotted lines/open symbols) versus the true estimated bias, shown for combinations of the biserial correlation $Corr(U, X) = \rho_{ux}$ (rows) and the selection mechanism (columns), for E[Y] = 0.5. Grey dashed line is equality (index = estimated bias). Results are medians across 1000 simulated data sets for each scenario.



Supplemental Figure 3: Coverage of [MUBP(0), MUBP(1)] and [SMUB(0), SMUB(1)] ML/MML intervals, and Bayesian credible intervals ("Bayes"), shown as a function of the true estimated bias (x-axis), selection mechanism and estimation method (columns), proxy strength (rows), and E[Y] (shape). Coverages are estimated from 1000 simulated data sets.