Technical Appendix to Analyzing a Randomized Cancer Prevention Trial with a Missing Binary Outcome, an Auxilliary Variable, and All-or-None Compliance

Stuart G. Baker

APPENDIX A: CALCULATIONS FOR MODEL A

ML estimates

The data are $\{n_{xay}, m_{xa}\}$. Let $N_x = n_{x++} + m_{x+}$, $h_{xay} = n_{xay}/N_x$ and $g_{xa} = m_{xa}/N_x$. To obtain ML estimates for a saturated model, we set observed proportions equal to their expected values:

$$\beta_{y|x} f_{a|xy} q_{ax} = h_{xay}, \tag{A.1}$$

$$\Sigma_y \beta_{y|x} f_{a|xy} p_{ax} = g_{xa.} \tag{A.2}$$

Summing (A.1) over y and adding to (A.2) gives

$$\Sigma_y \beta_{y|x} f_{a|xy} = h_{xa+} + g_{xa}. \tag{A.3}$$

Dividing (A.2) by the (A.3) gives $p_{ax} = g_{xa}/(h_{xa+} + g_{xa})$ which we substitute into (A.1) to obtain

$$\beta_{y|x} f_{a|xy} = h_{xay}(h_{xa+} + g_{xa})/h_{xa+} = h_{xay} + g_{xay}, \tag{A.4}$$

where $g_{xay} = h_{xay} g_{xa}/h_{xa+}$. Summing both sides of (A.4) over a gives $\widehat{\beta}^{\mathrm{A}}_{y|x} = h_{x+y} + g_{x+y}$.

We can also obtain the ML estimate for $\beta_{y|x}$ by applying the substitution principle to the identity $\Pr(Y = y|X = x) = \Sigma_a \Pr(Y = y, A = a|x) = \Sigma_a \Pr(Y = y, A = a, R = 1|x)/\Pr(R = 1|x, a, y)$. For the numerator we substitute $n_{xay}/(n_{x++} + m_{x+})$. For the denominator we invoke the missing-data mechanism and substitute $n_{xa+}/(n_{xa+} + m_{xa})$.

Asymptotic Variance

The asymptotic variances are $var(D\widehat{I}F^{\mathrm{A}}) = var(\widehat{\beta}^{\mathrm{A}}_{1|0}) + var(\widehat{\beta}^{\mathrm{A}}_{1|1})$ and $var(L\widehat{R}R^{\mathrm{A}}) = var(\widehat{\beta}^{\mathrm{A}}_{1|0})/(\widehat{\beta}^{\mathrm{A}}_{1|0})^2 + var(\widehat{\beta}^{\mathrm{A}}_{1|1})/(\widehat{\beta}^{\mathrm{A}}_{1|1})^2$. Using the MP transformation (Baker, 1994a),

$$var(\widehat{\beta}^{\mathrm{A}}_{1|x}) = \sum_{x=0}^{1}\sum_{a=0}^{1}\sum_{y=0}^{1} n_{xay}\left(\frac{\partial\widehat{\beta}^{\mathrm{A}}_{1|x}}{\partial n_{xay}}\right)^2 + \sum_{x=0}^{1}\sum_{a=0}^{1} m_{xa}\left(\frac{\partial\widehat{\beta}^{\mathrm{A}}_{1|x}}{\partial m_{xa}}\right)^2$$

$$= \sum_{x=0}^{1}\sum_{a=0}^{1}\Big\{ n_{xa0}\left(\tfrac{m_{xa1}}{n_{xa+}} + \widehat{\beta}^{\mathrm{A}}_{1|x}\right)^2 + n_{xa1}\left(-1 - \tfrac{m_{xa0}}{n_{xa+}} + \widehat{\beta}^{\mathrm{A}}_{1|x}\right)^2$$

$$+ m_{xa}\left(-\tfrac{n_{xa1}}{n_{xa+}} + \widehat{\beta}^{\mathrm{A}}_{1|x}\right)^2\Big\} /(N_x)^2.$$

APPENDIX B: CALCULATIONS FOR MODEL AM

ML estimates

The data are $\{n_{xay}, m_{xa}, l_{xy}, o_x\}$. Let $M_x = n_{x++} + m_{x+} + l_{x+} + o_{x+}$. To obtain ML estimates, we set observed proportions equal to their expected values:

$$\beta_{y|x} f_{a|xy} q_x q^*_{xy} q_{xa} = n_{xay}/M_{x,} \quad \text{(B.1)}$$

$$\beta_{y|x} f_{a|xy} q_x q^*_{xy} p_{xa} = m_{xa}/M_{x,} \quad \text{(B.2)}$$

$$\beta_{y|x} q_x p^*_{xy} = l_{xy}/M_{x,} \quad \text{(B.3)}$$

$$p_x = o_x/M_{x.} \quad \text{(B.4)}$$

Let $N_x = n_{x++} + m_{x+} + l_{x+}$ and define $h_{xay} = n_{xay}/N_{x,}$ $g_{xa} = m_{xa}/N_x$and $j_{xy} = l_{xy}/N_x$.Substituting (B.4) into (B.1), (B.2), and (B.3) gives

$$\beta_{y|x} f_{a|xy} q^*_{xy} q_{xa} = h_{xay}, \quad \text{(B.5)}$$

$$\Sigma_y \beta_{y|x} f_{a|xy} q^*_{xy} p_{xa} = g_{xa}, \quad \text{(B.6)}$$

$$\beta_{y|x} p^*_{xy} = j_{xy}. \quad \text{(B.7)}$$

Summing (B.5) over y and adding it to (B.6) gives

$$\Sigma_y \beta_{y|x} f_{a|xy} q^*_{xy} = h_{xa+} + g_{xa}. \quad \text{(B.8)}$$

Dividing (B.6) by (B.8) and solving gives $p_{xa} = g_{xa}/(h_{xa+} + g_{xa})$ and thus $q_{xa} = h_{xa+}/(h_{xa+} + g_{xa})$. Substituting into (B.5) and summing over a gives

$$\beta_{y|x} q^*_{xy} = \Sigma_a h_{xay}/\, q_{xa} \;= (h_{x+y} +\; g_{x+y}), \qquad \text{(B.9)}$$

where $g_{xay} =\; h_{xay}\, g_{xa}/h_{xa+}$. Adding (B.7) and (B.9) gives $\widehat{\beta}^{\text{AM}}_{y|x} = h_{x+y} \; + g_{x+y} + j_{xy}$.

We can also obtain the ML estimate for $\beta_{y|x}$ by applying the substitution principle to the identity $\Pr(Y = y|\, x) \;= \Sigma_a \Pr(A = a, Y = y|\, x) \;=\; \Sigma_a \Pr(A = a,\; Y = y,\; R_A = 1,$ $R_Y = 1|\; x) \;/\; \Pr(R_A = 1, R_Y = 1|x, a, y)$. For the numerator we substitute n_{xay} / $(n_{x++} \;+\; m_{x+} \;+\; l_{x+} \;+\; o_x)$. Under the missing-data mechanism, we write the denominator as $\Pr(R_{AY} = 1|x) \;\times \Pr(\; R_A = 1 \mid R_{AY} = 1,\; x, y) \times \Pr(R_Y = 1$ $|R_A = 1, x, a)$. For the first factor we substitute $(n_{x++} \;+\; m_{x+} \;+\; l_{x+}\,)/\,(n_{x++} \;+\; m_{x+}$ $+\; l_{x+} \;+\; o_x)$. For the second, nonignorable, factor we use the imputed counts and substitute $(n_{x+\text{y}} +\; m_{x+y})\;/(n_{x+y} + m_{x+y} + l_{xy})$. For the third factor we substitute $n_{xa+}/(n_{xa+}+ m_{xa})$.

Asymptotic variances

The asymptotic variances are $var(D\widehat{I}F^{\text{AM}}) = var(\widehat{\beta}^{\text{AM}}_{1|0}) + var(\widehat{\beta}^{\text{AM}}_{1|1})$ and $var(L\widehat{R}R^{\text{AM}}) =\; var(\widehat{\beta}^{\text{AM}}_{1|0})/(\widehat{\beta}^{\text{AM}}_{1|0})^2 \;+\;\; var(\widehat{\beta}^{\text{AM}}_{1|1})/(\widehat{\beta}^{\text{AM}}_{1|1})^2$. Using the MP transformation (Baker, 1994a),

$$var(\widehat{\beta}^{\text{AM}}_{1|x}) = \sum_{x=0}^{1}\sum_{a=0}^{1}\;\sum_{y=0}^{1} n_{xay}\left(\frac{\partial\widehat{\beta}^{\text{AM}}_{1|x}}{\partial n_{xay}}\right)^2 + \sum_{x=0}^{1}\sum_{a=0}^{1} m_{xa}\left(\frac{\partial\widehat{\beta}^{\text{AM}}_{1|x}}{\partial m_{xa}}\right)^2 + \sum_{x=0}^{1}\sum_{y=0}^{1} l_{xy}\left(\frac{\partial\widehat{\beta}^{\text{AM}}_{1|x}}{\partial l_{xy}}\right)^2$$

$$= 1/(N_x)^2 \quad \{n_{x00}\,(j_{x1} + e_{x0} r_{x0} + u_x)^2$$

$$+\, n_{x01}\,(j_{x0} \;- 1 - \eta_{x0} + e_{x0} r_{x0} + u_x)^2 \quad + n_{x10}\,(j_{x1} + e_{x1} r_{x1} + u_x)^2$$

$$+\; n_{x11}\,(j_{x0} - 1 - e_{x1} \quad + e_{x1} r_{x1} + u_x)^2 \quad + m_{x0}\,(j_{x1} \;- r_{x1} + u_x)^2$$

$$+\, m_{x1}\,(j_{x1} \;- r_{x1} + b_x)^2 \quad + l_{x0}(j_{x1} + u_x)^2 \;+\; l_{x1}(j_{x0} \;- 1 + u_x)^2\,\},$$

where $r_{xa} = n_{xa1}/n_{xa+}$, $e_{xa} \;= m_{xa}/n_{xa+}$, and $u_x = \Sigma_a n_{xa1}(1 + e_{xa})\,/N_x$.

APPENDIX C: CALCULATIONS FOR MODEL C

ML estimates

The data are $\{n_{xdy}, m_{xd}\}$. Let $N_x = n_{x++} + m_{x+}$ and $h_{xdy} = n_{xdy}/N_x$. To obtain ML estimates, we set observed proportions equal to their expected values:

$$\beta_{y|N}\, w_N\, q_N + \beta_{y|C0} w_C\, q_C = h_{00y}, \tag{C.1}$$

$$\beta_{y|A}\, w_A q_A = h_{01y}, \tag{C.2}$$

$$\beta_{y|N}\, w_N q_N = h_{10y}, \tag{C.3}$$

$$\beta_{y|C1}\, w_C\, q_C + \beta_{y|A}\, w_A q_A = h_{11y}\,. \tag{C.4}$$

To estimate $\beta_{y|C0}$, we subtract (C.2) from (C.4) to obtain

$$\beta_{y|C0} w_C q_C = h_{00y} - h_{10y}\,. \tag{C.5}$$

Dividing (C.5) its sum over y gives $\widehat{\beta}^{\mathrm{C}}_{y|C0} = (h_{00y} - h_{10y}) \,/\, (h_{00+} - h_{10+})$. We similarly derive $\widehat{\beta}^{\mathrm{C}}_{y|C1} = (h_{11y} - h_{01y}) \,/\, (h_{11+} - h_{01+})$.

We can also obtain the ML estimate for $\beta_{y|C0}$ by applying the substitution principle to the identity $\Pr(Y = y|C, D = 0) = \Pr(Y = y, R = 1\, |C, D = 0)/\Pr(R = 1|C, D = 0, y)$. Multiplying numerator and denominator by $\Pr(D = 0|C)$ gives $\Pr(Y = y, R = 1,\, D = 0\, |C)/\, pr(R = 1|D = 0, C, y)$. For the numerator, we substitute $h_{00y} - h_{10y}$; for the denominator, we assume missing does not depend on outcome and substitute $h_{00+} - h_{10+}$. We can similarly obtain the ML estimate for $\beta_{y|C1}$.

Asymptotic variance

The asymptotic variances are $var(D\widehat{I}F^{\mathrm{C}}) = var(\widehat{\beta}^{\mathrm{C}}_{1|C0}) + var(\widehat{\beta}^{\mathrm{C}}_{1|C0}) - 2\, cov(\widehat{\beta}^{\mathrm{C}}_{1|C0}, \widehat{\beta}^{\mathrm{C}}_{1|C1})$ and $var(L\widehat{R}R^{\mathrm{C}}) = var(\widehat{\beta}^{\mathrm{C}}_{1|C1})/(\widehat{\beta}^{\mathrm{C}}_{1|C1})^2 + var(\widehat{\beta}^{\mathrm{C}}_{1|C0})/(\widehat{\beta}^{\mathrm{C}}_{1|C0})^2 - 2\, cov(\widehat{\beta}^{\mathrm{C}}_{1|C0}, \widehat{\beta}^{\mathrm{C}}_{1|C1})/(\widehat{\beta}^{\mathrm{C}}_{1C0}, \widehat{\beta}^{\mathrm{C}}_{1C1})$. Based on a multinomial distribution,

$$var(\widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}d}) = \sum_{y=0}^{1} \left(\frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}d}}{\partial h_{11y}} \right)^2 \sum_x \frac{h_{x1y}(1-h_{x1y})}{N_x} - 2 \left(\frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}d}}{\partial h_{111}} \frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}d}}{\partial h_{110}} \right) \sum_x \frac{h_{x11} h_{x10}}{N_x}$$ and

$$cov(\widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}0}, \widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}1}) = -2 \sum_{\{d,y,d',y')\} \in E} \left(\frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}d}}{\partial h_{1dy}} \frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|\mathrm{C}d}}{\partial h_{1d'y'}} \right) \sum_x \frac{h_{xdy} h_{xd'y'}}{N_x},$$ where

$$E = \{\{1,1,0,1\}, \{1,1,0,0\}, \{0,1,1,0\}, \{1,0,0,0\}\},$$

$$\frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|c1}}{\partial h_{x11}} = (-1)^{(1-x)} \frac{1-\widehat{\beta}^{\mathrm{C}}_{1|c1}}{h_{11+}-h_{01+}}, \qquad \frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|c1}}{\partial h_{x10}} = (-1)^{(1-x)} \frac{-\widehat{\beta}^{\mathrm{C}}_{1|c1}}{h_{11+}-h_{01+}},$$

$$\frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|c0}}{\partial h_{x11}} = (-1)^{x} \frac{1-\widehat{\beta}^{\mathrm{C}}_{1|c0}}{h_{00+}-h_{10+}}, \quad \frac{\partial \widehat{\beta}^{\mathrm{C}}_{1|c0}}{\partial h_{x10}} = (-1)^{x} \frac{-\widehat{\beta}^{\mathrm{C}}_{1|c0}}{h_{00+}-h_{10+}}.$$

APPENDIX D: CALCULATIONS FOR MODEL AC

ML estimates

The data are $\{n_{xday}, m_{xda}\}$. Let $N_x = n_{x+++} + m_{x++}$, $h_{xday} = n_{xday}/N_x$ and $g_{xda} = m_{xda}/N_x$. Because the model is saturated, to obtain ML estimates we can set observed proportions equal to their expected values:

$$\beta_{y|N}\, w_N f_{a|Ny}\, q_{aN} + \beta_{y|C0} w_C f_{a|C0y}\, q_{aC} = h_{00ay}, \tag{D.1}$$

$$\beta_{y|A}\, w_A f_{a|Ay} q_{aA} = h_{01ay}, \tag{D.2}$$

$$\beta_{y|N}\, w_N f_{a|Ny}\, q_{aN} = h_{10ay}, \tag{D.3}$$

$$\beta_{y|C1}\, w_C f_{a|C1y}\, q_{aC} + \beta_{y|A}\, w_A f_{a|A0} q_{aA} = h_{11ay}, \tag{D.4}$$

$$\Sigma_y \beta_{y|N}\, w_N f_{a|Ny} p_{aN} + \beta_{y|C0} w_C f_{a|C0y}\, p_{aC} = g_{00a}, \tag{D.5}$$

$$\Sigma_y \beta_{y|A}\, w_A f_{a|Ay} p_{aA} = g_{01a}, \tag{D.6}$$

$$\Sigma_y \beta_{y|N}\, w_N f_{a|Ny} p_{aN} = g_{10a}, \tag{D.7}$$

$$\Sigma_y \beta_{y|C1}\, w_C f_{a|C1y}\, p_{aC} + \beta_{y|A}\, w_A f_{a|Ay}\, p_{aA} = g_{11a}. \tag{D.8}$$

We derive the estimate of $\beta_{y|C0}$. The derivation of the estimate of $\beta_{y|C1}$ is similar. Subtracting (D.3) from (D.1) gives

$$\beta_{y|C0} f_{a|C0y} q_{aC0} w_C = h_{00ay} - h_{10ay}. \tag{D.9}$$

Dividing both sides of (D.9) by $q_{aC0} w_C$ and summing over a gives

$$\widehat{\beta}^{\mathrm{AC}}_{y|C0} = \Sigma_a (h_{00ay} - h_{10ay}) / (\widehat{q}_{aC0} \widehat{w}_C). \tag{D.10}$$

To estimate q_{C0}, we first sum (D.9) over y to obtain

$$(\Sigma_y \beta_{y|C0}\, f_{a|C0y})\ q_{aC0} w_C\ = h_{00a+} -\ h_{10a+}, \tag{D.11}$$

and we subtract (D.7) from (D.5) to obtain

$$(\Sigma_y \beta_{y|C0}\, f_{a|C0y})\ p_{aC0} w_C = g_{00a} - g_{10a}. \tag{D.12}$$

Adding (D.11) and (D.12) gives

$$(\Sigma_y \beta_{y|C0} f_{a|C0y}) w_C\ = (h_{00a+} + g_{00a+}) - (h_{01a+} + g_{01a+}). \tag{D.13}$$

Dividing (D.11) by (D.13) gives $\widehat{q}_{aC0}$. Summing (D.1) to (D.8) over a and y and rearranging terms gives $\widehat{w}_C$.

We can also obtain the ML estimate for $\beta_{y|C0}$ by applying the substitution principle to the identity $\Pr(Y = y|\ C,\ D = 0)\ =\ \Sigma_a \Pr(A = a, Y = y, R = 1|C, D = 0)$ $pr(D = 0|C)$ / $[\Pr(R = 1|a, C, D = 0, y) \Pr(D = 0|C)]$. For the numerator we substitute $h_{00ay} - h_{10ay}$. For the first factor in the denominator, we invoke the missing-data mechanism and substitute $\widehat{q}_{aC0}$. For the second factor in the denominator, we substitute $\widehat{w}_C$. We can similarly obtain the ML estimate for $\beta_{y|C1}$.

Asymptotic variance

Using the MP transformation (Baker, 1994a), the asymptotic variance and covariance are

$$var(\ \widehat{\beta}^{\mathrm{AC}}_{1|Cd}) = \sum_x \sum_d \sum_a \sum_y n_{xady} \left(\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{1|Cd}}{\partial n_{xday}}\right)^2 + \sum_x \sum_d \sum_a m_{xad} \left(\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{1|Cd}}{\partial m_{xda}}\right)^2 \text{ and}$$

$$cov(\ \widehat{\beta}^{\mathrm{AC}}_{1|C0}, \widehat{\beta}^{\mathrm{AC}}_{1|C1}\ ) = \sum_x \sum_d \sum_a \sum_y n_{xday} \left(\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{1|C0}}{\partial n_{xday}}\right) \left(\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{1|C1}}{\partial n_{xday}}\right)$$

$$+ \sum_x \sum_d \sum_a m_{xda} \left(\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{1|C0}}{\partial m_{xda}}\right) \left(\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{1|C1}}{\partial m_{xday}}\right).$$

When $N_0 = N_1 \equiv N,$ after some algebra the derivatives simplify to

$$\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{y|C0}}{\partial n_{00a0}} = -s_{00} - u_{0a} \qquad \frac{\partial \widehat{\beta}^{\mathrm{AC}}_{y|C0}}{\partial n_{10a0}} = -s_{10} - t_{00} + u_{0a}$$

$$\frac{\partial \widehat{\beta}^{\mathrm{AC}}_{y|C0}}{\partial n_{001a}} = -s_{00} - u_{0\mathrm{a}} + v_{00a} \qquad \frac{\partial \widehat{\beta}^{\mathrm{AC}}_{y|C0}}{\partial n_{101a}} = -s_{10} - t_{00} + u_{0a} - v_{10a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial n_{01a0}} = -s_{00} - t_{00}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial n_{01a1}} = -s_{00} - t_{00}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial m_{00a}} = -s_{00} + e_{0a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial m_{01a}} = -s_{00} - t_{00}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{00a0}} = s_{01}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{001a}} = s_{01}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{01a0}} = s_{01} + t_{01} + u_{1a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{01a1}} = s_{01} + t_{01} + u_{1a} - v_{01a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial m_{00a}} = s_{01}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial m_{01a}} = s_{01} + t_{01} - e_{1a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial n_{0110}} = -s_{10}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial n_{11a0}} = -s_{10}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial m_{10a}} = -s_{10} - t_{10} - e_{0a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C0}}{\partial m_{11a}} = -s_{10}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{10a0}} = s_{11} + t_{11}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{101a}} = s_{11} + t_{11}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{0110}} = s_{11} - u_{1a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial n_{11a0}} = s_{11} - u_{1a} + v_{11a}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial m_{10a}} = s_{11} + t_{11}$$

$$\frac{\partial\hat{\beta}^{AC}_{y|C1}}{\partial m_{11a}} = s_{11} + e_{1a}$$

where

$$s_{0d} = \sum_a \frac{n_{0da1}}{N^2\widehat{q}_{aCd}\widehat{w}_C} - t_{00}\widehat{w}_A \qquad s_{1d} = \sum_a \frac{n_{1d1a}}{N^2\widehat{q}_{aCd}\widehat{w}_C} - t_{1d}\widehat{w}_N$$

$$t_{x0} = -\frac{\hat{\beta}^{AC}_{y|C0}}{N\,\widehat{w}_C} \qquad t_{x1} = \frac{\hat{\beta}^{AC}_{y|C1}}{N\,\widehat{w}_C}$$

$$e_{0a} = \frac{h_{00a1} - h_{10a1}}{(n_{00a+} - n_{10a+})\widehat{w}_C} \qquad e_{1a} = \frac{h_{11a1} - h_{01a1}}{(n_{11a+} - n_{01a+})\,\widehat{w}_C}$$

$$u_{da} = e_{da}\left(\frac{1}{\widehat{q}_{aCd}} - 1\right) \qquad v_{xd\mathrm{a}} = \frac{1}{N\widehat{q}_{aCd}\widehat{w}_C}$$

APPENDIX E: CALCULATION OF ASYMPTOTIC PERFORMANCE

We analytically approximate the asymptotic performances and check the results via simulation. As an example, consider the asymptotic performance of $D\widehat{I}F^{\mathrm{B}}$ under distribution C. Similar calculations apply to other distributions and to $L\widehat{R}R^{\mathrm{B}}$.

As a preliminary step, we compute the asymptotic variance of $D\widehat{I}F^{\text{B}}$ under distribution C. Under distribution C the data are $\{n_{xdy}, m_{xd}\}$ and under distribution B, the data are $\{n_{xy} = n_{x+y}\}$. Using the MP-transformation (Baker, 1994), the asymptotic variance of $D\widehat{I}F^{\text{B}}$ under distribution C is

$$var(D\widehat{I}F^{\text{B}}) = \sum_{x=0}^{1}\sum_{d=0}^{1}\sum_{y=0}^{1} var(n_{xdy})\left(\frac{\partial D\widehat{I}F^{\text{B}}}{\partial n_{xdy}}\right). \quad \text{(E.1)}$$

Because the model is saturated, $var(n_{xdy}) = n_{xdy}$. Therefore we can write (E.1) as

$$\begin{aligned} var(D\widehat{I}F^{\text{B}}) &= \sum_{x=0}^{1}\sum_{d=0}^{1}\sum_{y=0} n_{xdy}\left(\frac{\partial D\widehat{I}F^{\text{B}}}{\partial n_{x+y}}\right)\left(\frac{\partial n_{x+y}}{\partial n_{xdy}}\right) \\ &= \sum_{x=0}^{1}\sum_{y=0}^{1} n_{xy}\left(\frac{\partial D\widehat{I}F^{\text{B}}}{\partial n_{xy}}\right) \end{aligned} \quad \text{(E.2)}$$

Because $var(n_{xy}) = n_{xy}$ under distribution B, (E.2) is the asymptotic variance under distribution B computed via the MP transformation (Baker, 1995). Thus (E.2) is asymptotically equivalent to

$$var(D\widehat{I}F^{\text{B}}) = \sum_{x} \widehat{\beta}^{\text{B}}_{1|x}(1 - \widehat{\beta}^{\text{B}}_{1|x})/n_{x+}. \quad \text{(E.3)}$$

Two-sided type I error

We can approximate the true two-sided type I error under distribution C for a nominal 95% acceptance region testing if $D\widehat{I}F^{\text{B}}$ equals 0. Using (E.3), we compute se^{B}_{NUL}, the standard error of $D\widehat{I}F^{\text{B}}$ under distribution C. The lower and upper bounds of the 95% acceptance region for $D\widehat{I}F^{\text{B}}$ are then $L = -1.96\, se^{\text{B}}_{NUL}$ and $U = 1.96\, se^{\text{B}}_{NUL}$. To compute the true-two sided type I error for (L, U), let $D\widehat{I}F^{\text{B}}_{NUL}$ denote the estimate of DIF^{B} based on the expected counts from distribution C under the null hypothesis. Also let Φ denote the cumulative normal distribution with mean 0 and variance 1. The true two-sided type I error equals $1 - \Phi((U - D\widehat{I}F^{\text{B}}_{NUL})/\, se^{\text{B}}_{NUL}) + 1 - \Phi((\, D\widehat{I}F^{\text{B}}_{NUL} - L)/\, se^{\text{B}}_{NUL})$. To check via simulation, we compute se^{B}_{NUL}, L, U, and $D\widehat{I}F^{\text{B}}_{NUL}$ for each replication and count the fraction of times $D\widehat{I}F^{\text{B}}_{NUL}$ is outside (L, U).

Coverage

We compute an approximate true coverage for a nominal 95% confidence interval for $\widehat{DIF}^P$ Let $\widehat{DIF}^{\mathrm{B}}_{ALT}$ denote the estimate of DIF^{B}based on the expected counts from distribution C under the alternative hypothesis. Using (E.3) we compute se^{B}_{ALT} the standard error of $\widehat{DIF}^{\mathrm{B}}$ under distribution C. The nominal 95% confidence interval is (L^*, U^*) where $L^* = \widehat{DIF}^{\mathrm{B}}_{ALT} - 1.96\ se^{\mathrm{B}}_{ALT}$ and $U^* = \widehat{DIF}^{\mathrm{B}}_{ALT} + 1.96\ se^{\mathrm{B}}_{ALT}$. Because we generate data under distribution C, we know DIF^{C}. The p-values associated with the upper and lower bounds are $p_U = 1 - \Phi((U^* - DIF^{\mathrm{C}})\ /\ se^{\mathrm{B}}_{ALT})$ and $p_L = 1 - \Phi((DIF^{\mathrm{C}} - L^*)/\ se^{\mathrm{B}}_{ALT})$, respectively, so the true coverage is $1 - p_U - p_L$. In the simulations, we compute se^{B}_{ALT}, L^*, U^* for each replication and count the fraction of times the confidence interval (L^*, U^*) encloses DIF^{C}.

.