Appendix B

Let \( n_{ij0} \) denote the number in risk group \( i \) \((i = 1, 2, \ldots, k)\) and arm \( j \) \((j = 0 = \text{placebo}, j = 1 = \text{tamoxifen})\) without invasive breast cancer. Let \( n_{ij1} \) denote the total number in risk group \( i \) and arm \( j \) with invasive breast cancer. The likelihood kernel for the general formulation is

\[
L = \prod_i (1 - pr(\text{invasive breast cancer} \mid \text{placebo, group } i))^{n_{i00}}
\]
\[
\times pr(\text{invasive breast cancer} \mid \text{placebo, group } i)^{n_{i01}}
\]
\[
\times (1 - pr(\text{invasive breast cancer} \mid \text{tamoxifen, group } i))^{n_{i10}}
\]
\[
\times pr(\text{invasive breast cancer} \mid \text{tamoxifen, group } i)^{n_{i11}}.
\]

Recall that \( \pi_i = pr(\text{invasive breast cancer} \mid \text{placebo, group } i) \). Let \( \delta \) denote the absolute risk difference which is constant over risk groups. The kernel of the log-likelihood for the Constant RD model is

\[
L_{\text{Constant RD}}(\pi_i, \delta) = \sum_{i=1}^{k} n_{i00} \log (1 - \pi_i) + \sum_{i=1}^{k} n_{i01} \log (\pi_i)
\]
\[
+ \sum_{i=1}^{k} n_{i10} \log (1 - \pi_i + \delta) + \sum_{i=1}^{k} n_{i11} \log (\pi_i - \delta).
\]

Let \( \beta \) denote the relative risk which is constant over risk groups. The kernel of the log-likelihood for the Constant RR model is

\[
L_{\text{Constant RR}}(\pi_i, \beta) = \sum_{i=1}^{k} n_{i00} \log (1 - \pi_i) + \sum_{i=1}^{k} n_{i01} \log (\pi_i)
\]
\[
+ \sum_{i=1}^{k} n_{i10} \log (1 - \pi_i / \beta) + \sum_{i=1}^{k} n_{i11} \log (\pi_i / \beta).
\]
The above log-likelihoods were maximized using a Newton-Raphson algorithm with starting values of \( \pi_i = n_{i11}/n_{i1+} \), \( \delta = \Sigma_i(n_{i01}/n_{i0+} - n_{i11}/n_{i1+})/k \), and \( \beta = \Sigma_i((n_{i01}/n_{i0+}) / (n_{i11}/n_{i1+}))/k \), where "+" indicates summation over the indicated subscript. Confidence intervals are based on the asymptotic variance computed via the observed information matrix.

For the full model the estimates are \( \hat{\delta}_i = n_{i01}/n_{i0+} - n_{i11}/n_{i1+} \) and \( \hat{\beta}_i = (n_{i01}/n_{i0+}) /
(n_{i11}/n_{i1+}) \). Confidence intervals are based on the asymptotic variance for binomial distributions. The maximized log-likelihood for both *Varying RD* and *Varying RR* models is

\[
L_{VaryingRD}(\widehat{\pi}_i, \widehat{\delta}_i) = L_{VaryingRR}(\widehat{\pi}_i, \widehat{\beta}_i) = \\
\sum_{i=1}^{k} n_{i00} \log(n_{i00}/n_{i0+}) + \sum_{i=1}^{k} n_{i01} \log(n_{i01}/n_{i0+}) \\
+ \sum_{i=1}^{k} n_{i10} \log(n_{i10}/n_{i1+}) + \sum_{i=1}^{k} n_{i11} \log(n_{i11}/n_{i1+}).
\]

Based on an asymptotic chi-squared distribution, p-values for comparing models are computed for \( 2(L_{VaryingRD}(\widehat{\pi}_i, \widehat{\delta}_i) - L_{ConstantRD}(\widehat{\pi}_i, \widehat{\delta}_i)) \) and \( 2(L_{VaryingRR}(\widehat{\pi}_i, \widehat{\beta}_i) - L_{ConstantRD}(\widehat{\pi}_i, \widehat{\beta}_i)) \).