Appendix 1: The Formulas

For confounder adjustment in an unmatched design, to extend the formulas in [1], page 305, from one to F controls per case, we need only multiply the parameter Σ' in [1] by F.

$$\text{We define } \Sigma' = F[PR_E \sum_{i=1}^{K} p_{1i}R_{Ci} + (1-P)\sum_{i=1}^{K} p_{2i}R_{Ci}] \text{ , and}$$

$$V_A = (\sum 1/V_{iA})^{-1} \quad , \quad V_{iA} = \frac{1}{Pp_{1i}} + \frac{1}{(1-P)p_{2i}} + \Sigma' \left[\frac{1}{Pp_{1i}R_{Ci}R_E} + \frac{1}{(1-P)p_{2i}R_{Ci}}\right] ,$$

$$V_N = (\sum 1/V_{iN})^{-1} \quad , \quad V_{iN} = T_i^3/W_{1i}W_{2i}W_{3i}W_{4i}$$

(we divide out the factor n shown in [1]), where the proportion in confounder level i

$$\text{for a control is } \; W_{1i} = Pp_{1i} + (1-P)p_{2i} \; ,$$

$$\text{for a case is } \; W_{2i} = [Pp_{1i}R_E + (1-P)p_{2i}]R_{Ci}/\Sigma' \; ,$$

$$\text{and exposed is } \; W_{3i} = Pp_{1i}(1 + R_{Ci}R_E/\Sigma') \; ,$$

$$\text{and unexposed is } \; W_{4i} = (1-P)p_{2i}(1 + R_{Ci}/\Sigma') \; ; \quad \text{and}$$

$$T_i = W_{1i} + W_{2i} = W_{3i} + W_{4i} \; .$$

To compute the sample size, we invert formula (7.15) of [1], to obtain the one-sided ($S = 1$) solution

$$n = [(z_\alpha\sqrt{V_N} + z_{1-\beta}\sqrt{V_A})/\log(R_E)]^2 \; , \tag{1}$$

and the control sample size is the first integer greater than n. The case sample size is the first integer greater than n/F. The terms z_α and $z_{1-\beta}$ are the usual Standard Normal deviates. We replace z_α with $z_{\alpha/2}$ for the two-sided solution (when $S = 2$). Four common values of α and two values of β are supplied by the program.

The unadjusted formulas are shown on page 305 of [1], but it is easy to show that these are the same as the results obtained when we specify $K = 1$ in the preceding formulas. Note that $p_{11} = 1$ and $p_{21} = 1$ in that case. In the computer program, therefore, we use the same lines of computation for the adjusted and unadjusted results.

For confounder adjustment in a strata-matched design, The definition of Σ' is the same as above. The other items are re-defined as follows:

$$V_{iA} = \Sigma'\left\{\left[\frac{1}{Pp_{1i}} + \frac{1}{(1-P)p_{2i}}\right]\frac{[Pp_{1i} + (1-P)p_{2i}]}{[Pp_{1i}R_E + (1-P)p_{2i}]R_{Ci}F} + \frac{1}{Pp_{1i}R_{Ci}R_E} + \frac{1}{(1-P)p_{2i}R_{Ci}}\right\} ,$$

$$V_{iN} = T_i^3/W_{1i}W_{2i}W_{3i}W_{4i} = (1+F)^3W_{2i}/FW_{3i}W_{4i} \text{ , with}$$

$$W_{1i} = FW_{2i} \; ,$$

$$W_{2i} = [Pp_{1i}R_E + (1-P)p_{2i}]R_{Ci}/\Sigma' \text{ , as before,}$$
$$W_{3i} = Pp_{1i}R_{Ci}R_E/\Sigma' + FW_{2i}Pp_{1i}/[Pp_{1i} + (1-P)p_{2i}] \text{ ,}$$
$$W_{4i} = (1-P)p_{2i}R_{Ci}/\Sigma' + FW_{2i}(1-P)p_{2i}/(Pp_{1i} + (1-P)p_{2i}) \text{ ,}$$
$$\text{and } T_i = W_{1i} + W_{2i} = (1+F)W_{2i} = W_{3i} + W_{4i} \text{ .}$$

Again, we use $V_A = (\sum 1/V_{iA})^{-1}$ and $V_N = (\sum 1/V_{iN})^{-1}$, which are the approximations due to Woolf [2].

To compute the sample size, we again use formula (1).

References

[1] NE Breslow, NE Day: *Statistical Methods in Cancer Research, Vol. 2: The Design and Analysis of Cohort Studies*, IARC Scientific Publications No. 82, International Agency of Research on Cancer, Lyon, France, 1987, Sections 7.8-7.9, 305-306.

[2] B Woolf: **On estimating the relationship between blood group and disease**. *Annals of Human Genetics*, 1955 **19**, pp. 251-253.