Analysis of paired, screen-positive designs

Mark Clements and Martin Eklund

2015-02-20

1 Model extension to Pepe and Alonzo (2001)

Pepe and Alonzo (2001) show how to fit regression models to paired, screen-positive designs. They use marginal models fitted to all of the data, including those with missing disease status. We propose a simple extension to their framework, where the model is fitted to those individuals with known disease status.

Following the notation used by Pepe and Alonzo (2001), we consider 0/1 tests Y_A and Y_B, with disease status D and covariates Z. Pepe and Alonzo define $\alpha Z = \log P(D=1, Y_B=1|Z)$ and $\alpha Z + \beta Z = \log P(D=1, Y_A=1|Z)$, and use $T=1$ for test A and $T=0$ for test B with Y for the test value.

Similar, we define an indicator $S = I(Y_A = 1 \text{ or } Y_B = 1)$ for whether one or either of the tests is positive, such that the disease status is known. Moreover, let $\alpha^* Z = \log P(D=1, Y_B=1|Z, S=1)$ and $\alpha^* Z + \beta^* Z = \log P(D=1, Y_A=1|Z, S=1)$, where we have conditioned on having one or more positive tests. Then, closely following the argument in Pepe and Alonzo (2001),

$$\begin{aligned}
\beta^* Z &= \log \frac{P(D=1, Y_A=1|Z, S=1)}{P(D=1, Y_A=1|Z, S=1)} \\
&= \log \frac{P(Y=1|D=1, Z, T=1, S=1)P(D=1|Z, T=1, S=1)}{P(Y=1|D=1, Z, T=0, S=1)P(D=1|Z, T=0, S=1)} \\
&= \log \frac{P(Y_A=1|D=1, Z, S=1)}{P(Y_B=1|D=1, Z, S=1)}
\end{aligned}$$

where $P(D=1|Z, T=1, S=1) = P(D=1|Z, T=0, S=1)$ for paired data. Noting that

$$\begin{aligned}
P(Y=1|D=1, Z, T, S=1) &= \frac{P(Y=1|D=1, Z, T) - P(Y=1|D=1, Z, T, S=0)P(S=0)}{P(S=1)} \\
&= \frac{P(Y=1|D=1, Z, T)}{P(S=1)}
\end{aligned}$$

as $P(Y=1|D=1, Z, T, S=0) = 0$, because no positive tests have $S=0$, then

$$\begin{aligned}
\beta^* Z &= \log \frac{P(Y_A=1|D=1, Z, S=1)}{P(Y_B=1|D=1, Z, S=1)} \\
&= \log \frac{P(Y_A=1|D=1, Z)}{P(Y_B=1|D=1, Z)} \\
&= \beta Z = \text{rTPF}(Z)
\end{aligned}$$

In general, α^* does not equal α, as $P(D = 1, Y_B = 1|Z, S = 1)$ may be different to $P(D = 1, Y_B = 1|Z)$, and the intercept term also be shifted.

Care is further required in using a simplified model. Pepe and Alonzo (2001) fit a reduced PSA model by excluding the main effect due to race. However, [data-driven observation that requires a mathematical derivation], the estimated interaction effect for β^* is then a biased estimate of β.