Mantel-Haenszel odds ratio and Peto's log odds ratio under the general framework of combining CDs

Guang Yang, Pixu Shi, Minge Xie

Objective: We demonstrate that the widely used Mantel-Haenszel odds ratio is a special case of combining aCDs and Peto's log odds ratio can also be derived using combination of CD-like functions.

1. Model Setting:

There are K independent studies, i.e. K 2×2 tables as below:

	Event	NonEvent	Total
Experimental	X_i	$n_i - X_i$	n_i
Control	Y_i	$m_i - Y_i$	m_i
Total	t_i	$N_i - t_i$	N_i

- $X_i \sim \text{Binomial}(n_i, p_1)$, $Y_i \sim \text{Binomial}(m_i, p_2)$, $i = 1, \cdots, K$.
- $X_i | X_i + Y_i = t \sim$ Noncentral Hypergeometric Distribution:

$$P_\psi(X_i = x|t) = \binom{n_i}{x}\binom{m_i}{t-x}\psi^x / \sum_s \binom{n_i}{s}\binom{m_i}{t-s}\psi^s, \quad i = 1, \cdots, K$$

  where $\psi = \dfrac{p_1(1-p_2)}{p_2(1-p_1)}$.

- Objects of interest: odds ratio ψ, log odds ratio $\theta = \ln \psi$.

2. Mantel-Haenszel and Peto methods for combining odds ratio

- Mantel-Haenszel odds ratio:
  - Estimate for odds ratio of individual studies:

$$\hat{\psi}_i = \frac{R_i}{S_i} = \frac{X_i(m_i - Y_i)/N_i}{Y_i(n_i - X_i)/N_i} \tag{1}$$

$$\widehat{Var}(\hat{\psi}_i) = \hat{\psi}_i^2\left(\frac{P_i}{R_i} + \frac{Q_i}{S_i}\right) = \hat{\psi}_i^2\left(\frac{1}{X_i} + \frac{1}{n_i - X_i} + \frac{1}{Y_i} + \frac{1}{m_i - Y_i}\right) \tag{2}$$

    where $R_i = \dfrac{X_i(m_i - Y_i)}{N_i}$, $S_i = \dfrac{Y_i(n_i - X_i)}{N_i}$, $P_i = \dfrac{X_i + m_i - Y_i}{N_i}$, $Q_i = \dfrac{Y_i + n_i - X_i}{N_i}$.
    The $\widehat{Var}(\hat{\psi}_i)$ here is a widely used estimate for variance of odds ratio.

  - Estimate for the combined odds ratio:

$$\hat{\psi}_{MH} = \frac{\sum_i R_i}{\sum_i S_i} = \frac{\sum_i X_i(m_i - Y_i)/N_i}{\sum_i Y_i(n_i - X_i)/N_i} \tag{3}$$

$$\widehat{Var}(\hat{\psi}_{MH}) = \hat{\psi}_{MH}^2\left[\frac{\sum P_i R_i}{2(\sum_i R_i)^2} + \frac{\sum(P_i S_i + Q_i R_i)}{2(\sum_i R_i)(\sum_i S_i)} + \frac{\sum Q_i S_i}{2(\sum_i S_i)^2}\right] \tag{4}$$

    The $\widehat{Var}(\hat{\psi}_{MH})$ here is proposed by Robins, Breslow and Greenland(1986).

- Peto's odds ratio
  - Estimate for log odds ratio of individual studies:

$$\hat{\theta}_i = \frac{O_i - E_i}{V_i} \quad (5)$$

$$\widehat{Var}(\hat{\theta}_i) = \frac{1}{V_i} \quad (6)$$

    where $O_i = X_i$, $E_i = \frac{t_i n_i}{N_i}$, $V_i = \frac{t_i m_i n_i (N_i - t_i)}{N_i^2 (N_i - 1)}$.

  - Estimate for the combined log odds ratio:

$$\hat{\theta}_{Peto} = \frac{\sum_i O_i - \sum_i E_i}{\sum_i V_i} \quad (7)$$

$$\widehat{Var}(\hat{\theta}_{Peto}) = \frac{1}{\sum_i V_i} \quad (8)$$

3. Definition of cofidence distribution (CD) and combining CDs:

- A definition of CD can be found in Xie, Singh and Strawdermanl (2011):
  "A CD function $H(\cdot) = H(\mathbf{X}, \cdot)$ is a mapping from $\mathcal{X} \times \Theta \to [0,1]$ where, for each given sample $\mathbf{X} \in \mathcal{X}$, $H(\cdot)$ is a sample-dependent continuous cumulative distribution function on Θ. Also, we require that, when $\theta = \theta_0$ the true parameter value, $H(\theta_0) \equiv H(\mathbf{X}, \theta_0)$, as a function of the sample $\mathbf{X}$, follows the uniform distribution U[0,1]."

- Given independent CD functions $H_1(\theta), \cdots, H_K(\theta)$, one general recipe of combining them from Xie, Singh and Strawdermanl (2011) is:

$$H^{(c)}(\theta) = G_c\{g_c(H_1(\theta), \cdots, H_K(\theta))\}$$
$$g_c(u_1, \cdots, u_K) = \sum_{i=1}^{K} w_i F_0^{-1}(u_i) \quad (9)$$

  where $G_c(t) = P(g_c(U_1, \cdots, U_K) \leqslant t)$ and $U_1, \cdots, U_K$ are independent U[0,1] distribution.

4. Mantel-Haeszel odds ratio under framework of combining aCDs:

- Claim: $H_i(\psi) = \Phi(\frac{\psi - \hat{\psi}_i}{\widehat{Var}^{1/2}(\hat{\psi}_i)})$ is an aCD for ψ, where $\hat{\psi}_i$ and $\widehat{Var}^{1/2}(\hat{\psi}_i)$ are given by (1) and (2).
  Proof of the Claim:
  When $m_i, n_i \to \infty$ and $n_i/N_i \to l_i$, it is easy to get $\hat{\psi}_i \to \psi$ and $N_i \widehat{Var}(\hat{\psi}_i) \to \psi^2(\frac{1}{l_i p_1(1-p_1)} + \frac{1}{(1-l_i)p_2(1-p_2)})$. Using delta method,

$$\sqrt{N_i}(log(\hat{\psi}_i) - log(\psi)) \to N(0, \frac{1}{l_i p_1(1-p_1)} + \frac{1}{(1-l_i)p_2(1-p_2)})$$

$$\sqrt{N_i}(\hat{\psi}_i - \psi) \to N(0, \psi^2(\frac{1}{l_i p_1(1-p_1)} + \frac{1}{(1-l_i)p_2(1-p_2)}))$$

Therefore, $H_i(\psi) = \Phi(\frac{\psi - \hat{\psi}_i}{\widehat{Var}^{1/2}(\hat{\psi}_i)})$ is an aCD for ψ_i as $m_i, n_i \to \infty, n_i/N_i \to l_i$.

- Claim: $H_{MH}(\psi) = \Phi(\frac{\psi - \hat{\psi}_{MH}}{\widehat{Var}^{1/2}(\hat{\psi}_{MH})})$ is an aCD for ψ, where $\hat{\psi}_{MH}$ and $\widehat{Var}^{1/2}(\hat{\psi}_{MH})$ are given by (3) and (4).
  Proof of the Claim:
  When $m_i, n_i \to \infty, n_i/N_i \to l_i$ and $N_i/N_+ \to r_i$ where $N_+ = \sum_{i=1}^K N_i$, it is easy to get $\hat{\psi}_{MH} \to \psi$. And under the same limiting condition, Robins, Breslow and Greenland(1986) showed that $N_+\widehat{Var}(\hat{\psi}_{MH})$ is consistent for $N_+ Var^A(\hat{\psi}_{MH})$, where Var^A stands for asymptotic variance.
  Therefore, $H_{MH}(\psi) = \Phi(\frac{\psi - \hat{\psi}_{MH}}{\widehat{Var}^{1/2}(\hat{\psi}_{MH})})$ is an aCD for ψ as $m_i, n_i \to \infty, n_i/N_i \to l_i$ and $N_i/N_+ \to r_i$.

- Claim: $H_{MH}(\psi)$ can be get by combining $H_i(\psi)$.
  Proof of the Claim:
  Set $w_i = \frac{S_i\widehat{Var}^{1/2}(\hat{\psi}_i)}{(\sum_i S_i)\widehat{Var}^{1/2}(\hat{\psi}_{MH})}$. Using the combing recipe given by (9), if we choose weights to be $\frac{w_i}{\sum w_i^2}$ and $F_0(\cdot) = \Phi(\cdot)$, then $G_c(t) = \Phi(t)$ and a combined aCD for ψ is:

$$\begin{aligned} H^{(c)}(\psi) &= \Phi\left(\frac{1}{\sum_i w_i^2}\sum_i w_i \Phi^{-1}(H_i(\theta))\right) \\ &= \Phi\left(\frac{1}{\sum_i w_i^2}\sum_i \frac{S_i\widehat{Var}(\hat{\psi}_i)}{\sum_i S_i \widehat{Var}^{1/2}(\hat{\psi}_{MH})}\frac{\psi - \hat{\psi}_i}{\widehat{Var}^{1/2}(\hat{\psi}_i)}\right) \\ &= \Phi\left(\frac{\psi - \hat{\psi}_{MH}}{\widehat{Var}^{1/2}(\hat{\psi}_{MH})\sum_i w_i^2}\right) \end{aligned}$$

  Note that $\sum_i w_i^2 \neq 1$. However, $\sum_i w_i^2 \to 1$ as $m_i, n_i \to \infty, n_i/N_i \to l_i$ and $N_i/N_+ \to r_i$. Therefore, $H^{(c)}(\psi) \to H_{MH}(\psi)$.

5. Peto's log odds ratio and combining CD-like functions

- A CD-like function for individual log odds ratio is:

$$\widetilde{H}_i(\theta) = \Phi(\frac{\theta - \hat{\theta}_i}{\widehat{Var}^{1/2}(\hat{\theta}_i)}) = \Phi(\sqrt{V_i}\theta - \frac{O_i - E_i}{\sqrt{V_i}})$$

  Here we say it is a CD-like function because they are not exactly CD functions: the distribution of $H_i(\theta)$ and $H_{Peto}(\theta)$ is only uniformly distributed when $\theta = 0$. This is mainly because Peto's odds ratio and variance estimate are derived using Taylor expansion of likelihood at $\theta = 0$.

- A CD-like function for combined log odds ratio is:

$$H_{Peto}(\theta) = \Phi(\frac{\theta - \hat{\theta}_{Peto}}{\widehat{Var}^{1/2}(\hat{\theta}_{Peto})}) = \Phi(\sqrt{\sum_i V_i}\theta - \frac{\sum_i O_i - \sum_i E_i}{\sqrt{\sum_i V_i}})$$

- Claim: $H_{Peto}(\theta)$ can be get by combining $\widetilde{H}_i(\theta)$.
  Proof of the Claim:

  Using the combining recipe given by (9), if we choose weights to be $w_i = \sqrt{\frac{V_i}{\sum_i V_i}}$ and $F_0(\cdot) = \Phi(\cdot)$, then $\sum_i w_i^2 = 1$, $G_c(t) = \Phi(t)$ and:

$$\begin{aligned} H^{(c)}(\theta) &= \Phi(\sum_i w_i \Phi^{-1}(\widetilde{H}_i(\theta))) \\ &= \Phi(\sum_i \sqrt{\frac{V_i}{\sum_i V_i}}(\sqrt{V_i}\theta - \frac{O_i - E_i}{\sqrt{V_i}})) \\ &= \Phi(\sqrt{\sum_i V_i}\theta - \frac{\sum_i O_i - \sum_i E_i}{\sqrt{\sum_i V_i}}) \\ &= H_{Peto}(\theta) \end{aligned}$$

References

[1] N. Breslow. Odds Ratio Estimators When the Data are Sparse. *Biometrika*, 68, 73–84, 1981.

[2] N. Breslow and K. Liang. The variance of the Mantel-Haenszel Estimator. *Biometrics*, 38, 943–952, 1982.

[3] W. Hauck, The large sample variance of the Mantel-Haenszel Estimator of a common odds ratio. *Biometrics*, 35, 817–819, 1979.

[4] W. Hauck, S. Anderson and F. Leahy. Finite-Sample Properties of Some Old and Some New Estimators of a Common Odds Ratio from Multiple 2×2 Tables. *Journal of American Statistical Association*, 77, 145–152, 1982.

[5] P. McCullagh and J. Nelder. *Generalized Linear Models*, Chapman and Hall, 1983.

[6] N. Mantel and W. Haenszel. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748, 1959.

[7] A. Phillips and P. Holland. Estimators of the Variance of the Mantel-Haenszel Log-Odds-Ratio Estimate. *Biometrics*, 43, 425–431, 1987.

[8] J. Robins, N. Breslow and S. Greenland. Estimators of the Mantel-Haenszel Variance Consistent in Both Sparse Data and Large-Strata Limiting Models. *Biometrics*, 42, 311–323, 1986.

[9] S. Yusuf, R. Peto, J. Lewis, R. Colins and P. Sleight. Beta blockade during and after myocardial infarction-An overview of randomized trials. *Progress in Cardiovascular Disease*, 27, 335–371, 1985.

[10] M. Xie, K. Singh, and W. E. Strawderman. Confidence Distributions and a Unifying Framework for Meta-Analysis. *Journal of the American Statistical Association*, 106, 320–333, 2011.