Biostatistics (2018), **0**, 0, *pp.* 1–7
doi:10.1093/biostatistics/supplement˙SWD˙paper˙biostatistics˙revision˙v3

Supplementary material for "A Maximum Likelihood Approach to Power Calculations for Stepped Wedge Designs of Binary Outcomes"

Xin Zhou

Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

Xiaomei Liao

Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. Currently employed at AbbVie Inc., North Chicago, IL 60064, USA

Lauren M. Kunz

Deloitte, New York, NY 10112, USA

Sharon-Lise T. Normand

Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

Molin Wang, Donna Spiegelman*

Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

stdls@hsph.harvard.edu

*To whom correspondence should be addressed.

S1. DERIVATIVES OF LOG-LIKELIHOOD FUNCTIONS

When there are no time effects, by Leibniz's rule for differentiation with integration, the derivatives of the log-likelihood function (2.5) in the manuscript are, when $0 < \beta < 1$,

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \mu} = \frac{F_i(-\mu;\mu,\beta)e^{-\frac{\mu^2}{2\tau^2}} - F_i(1-\mu-\beta;\mu,\beta)e^{-\frac{(1-\mu-\beta)^2}{2\tau^2}} + \int_{-\mu}^{1-\mu-\beta} \frac{\partial}{\partial\mu}F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db}{\int_{-\mu}^{1-\mu-\beta} F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db} - \frac{e^{-\frac{\mu^2}{2\tau^2}} - e^{-\frac{(1-\mu-\beta)^2}{2\tau^2}}}{\int_{-\mu}^{1-\mu-\beta} e^{-\frac{b^2}{2\tau^2}}\,db}, \tag{S1.1}$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta} = \frac{-F_i(1-\mu-\beta;\mu,\beta)e^{-\frac{(1-\mu-\beta)^2}{2\tau^2}} + \int_{-\mu}^{1-\mu-\beta} \frac{\partial}{\partial\beta}F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db}{\int_{-\mu}^{1-\mu-\beta} F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db} + \frac{e^{-\frac{(1-\mu-\beta)^2}{2\tau^2}}}{\int_{-\mu}^{1-\mu-\beta} e^{-\frac{b^2}{2\tau^2}}\,db}, \tag{S1.2}$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial (\tau^2)} = \frac{\int_{-\mu}^{1-\mu-\beta} F_i(b;\mu,\beta)b^2e^{-\frac{b^2}{2\tau^2}}\,db}{2\tau^4\int_{-\mu}^{1-\mu-\beta} F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db} - \frac{\int_{-\mu}^{1-\mu-\beta} b^2e^{-\frac{b^2}{2\tau^2}}\,db}{2\tau^4\int_{-\mu}^{1-\mu-\beta} e^{-\frac{b^2}{2\tau^2}}\,db}, \tag{S1.3}$$

where the derivative $\frac{\partial}{\partial\mu}F_i(b;\mu,\beta) = F_i(b;\mu,\beta)\left(\frac{Z_{i01}}{\mu+b} - \frac{Z_{i00}}{1-(\mu+b)} + \frac{Z_{i11}}{\mu+\beta+b} - \frac{Z_{i10}}{1-(\mu+\beta+b)}\right)$ and the other derivative $\frac{\partial}{\partial\beta}F_i(b;\mu,\beta) = F_i(b;\mu,\beta)\left(\frac{Z_{i11}}{\mu+\beta+b} - \frac{Z_{i10}}{1-(\mu+\beta+b)}\right)$. Similarly, when $-1 < \beta < 0$, we can obtain another set of derivatives by the Leibniz's rule,

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \mu} = \frac{F_i(-\mu-\beta;\mu,\beta)e^{-\frac{(\mu+\beta)^2}{2\tau^2}} - F_i(1-\mu;\mu,\beta)e^{-\frac{(1-\mu)^2}{2\tau^2}} + \int_{-\mu-\beta}^{1-\mu} \frac{\partial}{\partial\mu}F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db}{\int_{-\mu-\beta}^{1-\mu} F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db} - \frac{e^{-\frac{(\mu+\beta)^2}{2\tau^2}} - e^{-\frac{(1-\mu)^2}{2\tau^2}}}{\int_{-\mu-\beta}^{1-\mu} e^{-\frac{b^2}{2\tau^2}}\,db}, \tag{S1.4}$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta} = \frac{F_i(-\mu-\beta;\mu,\beta)e^{-\frac{(\mu+\beta)^2}{2\tau^2}} + \int_{-\mu-\beta}^{1-\mu} \frac{\partial}{\partial\beta}F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db}{\int_{-\mu-\beta}^{1-\mu} F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db} + \frac{e^{-\frac{(\mu+\beta)^2}{2\tau^2}}}{\int_{-\mu-\beta}^{1-\mu} e^{-\frac{b^2}{2\tau^2}}\,db}, \tag{S1.5}$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial (\tau^2)} = \frac{\int_{-\mu-\beta}^{1-\mu} F_i(b;\mu,\beta)b^2e^{-\frac{b^2}{2\tau^2}}\,db}{2\tau^4\int_{-\mu-\beta}^{1-\mu} F_i(b;\mu,\beta)e^{-\frac{b^2}{2\tau^2}}\,db} - \frac{\int_{-\mu-\beta}^{1-\mu} b^2e^{-\frac{b^2}{2\tau^2}}\,db}{2\tau^4\int_{-\mu-\beta}^{1-\mu} e^{-\frac{b^2}{2\tau^2}}\,db}. \tag{S1.6}$$

Although the log-likelihood function (2.5) is not differentiable at $\beta = 0$, there is little interest in power calculations at $\beta = 0$. Nevertheless, if it were the case, a very small perturbation can be added to β to avoid this problem.

When there are time effects, the derivatives of the log-likelihood function (2.15) in the manuscript are as follows,

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \mu} = \frac{\left(\frac{\partial m(\boldsymbol{\theta})}{\partial \mu}\right) F_i(-m(\boldsymbol{\theta});\boldsymbol{\theta}) e^{-\frac{m(\boldsymbol{\theta})^2}{2\tau^2}} - \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \mu}\right) F_i(1-M(\boldsymbol{\theta});\boldsymbol{\theta}) e^{-\frac{(1-M(\boldsymbol{\theta}))^2}{2\tau^2}}}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} \tag{S1.7}$$

$$+\frac{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} \frac{\partial}{\partial \mu} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} - \frac{\left(\frac{\partial m(\boldsymbol{\theta})}{\partial \mu}\right) e^{-\frac{m(\boldsymbol{\theta})^2}{2\tau^2}} - \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \mu}\right) e^{-\frac{(1-M(\boldsymbol{\theta}))^2}{2\tau^2}}}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} e^{-\frac{b^2}{2\tau^2}}\, db},$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta} = \frac{\left(\frac{\partial m(\boldsymbol{\theta})}{\partial \beta}\right) F_i(-m(\boldsymbol{\theta});\boldsymbol{\theta}) e^{-\frac{m(\boldsymbol{\theta})^2}{2\tau^2}} - \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \beta}\right) F_i(1-M(\boldsymbol{\theta});\boldsymbol{\theta}) e^{-\frac{(1-M(\boldsymbol{\theta}))^2}{2\tau^2}}}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} \tag{S1.8}$$

$$+\frac{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} \frac{\partial}{\partial \beta} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} - \frac{\left(\frac{\partial m(\boldsymbol{\theta})}{\partial \beta}\right) e^{-\frac{m(\boldsymbol{\theta})^2}{2\tau^2}} - \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \beta}\right) e^{-\frac{(1-M(\boldsymbol{\theta}))^2}{2\tau^2}}}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} e^{-\frac{b^2}{2\tau^2}}\, db},$$

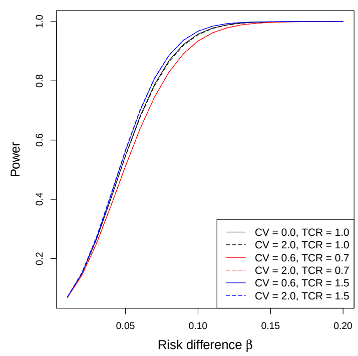
$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \gamma_j} = \frac{\left(\frac{\partial m(\boldsymbol{\theta})}{\partial \gamma_j}\right) F_i(-m(\boldsymbol{\theta});\boldsymbol{\theta}) e^{-\frac{m(\boldsymbol{\theta})^2}{2\tau^2}} - \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \gamma_j}\right) F_i(1-M(\boldsymbol{\theta});\boldsymbol{\theta}) e^{-\frac{(1-M(\boldsymbol{\theta}))^2}{2\tau^2}}}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} \tag{S1.9}$$

$$+\frac{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} \frac{\partial}{\partial \gamma_j} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} - \frac{\left(\frac{\partial m(\boldsymbol{\theta})}{\partial \gamma_j}\right) e^{-\frac{m(\boldsymbol{\theta})^2}{2\tau^2}} - \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \gamma_j}\right) e^{-\frac{(1-M(\boldsymbol{\theta}))^2}{2\tau^2}}}{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} e^{-\frac{b^2}{2\tau^2}}\, db}, \quad \text{for } j = 2, \cdots, J,$$

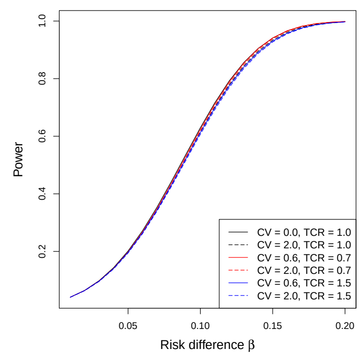
$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial (\tau^2)} = \frac{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) b^2 e^{-\frac{b^2}{2\tau^2}}\, db}{2\tau^4 \int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} F_i(b;\boldsymbol{\theta}) e^{-\frac{b^2}{2\tau^2}}\, db} - \frac{\int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} b^2 e^{-\frac{b^2}{2\tau^2}}\, db}{2\tau^4 \int_{-m(\boldsymbol{\theta})}^{1-M(\boldsymbol{\theta})} e^{-\frac{b^2}{2\tau^2}}\, db}. \tag{S1.10}$$

Next, we investigate the differentiability of the log-likelihood function (2.15). From (S1.7)$-$(S1.10), (2.15) is not differentiable when at least two elements of $\{\mu+\gamma_j, \mu+\beta+\gamma_j; j = 1, \cdots, J\}$ coincide at $m(\boldsymbol{\theta})$ or $M(\boldsymbol{\theta})$. It should be noted that the log-likelihood function (2.15) is not differentiable in a small number of places, such as when $\beta = 0$ or when $\gamma_j = \gamma_k$ for some j and k in $\{1, \cdots, J\}$. The first case is not of interest; for the second, steps for which the same time effect is to be assumed should be combined at the design stage.

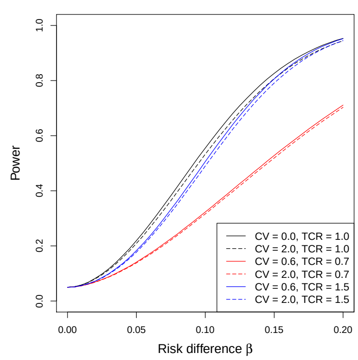
S2. Additional Figures and Tables



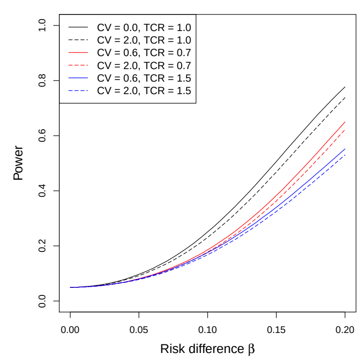
(a) $\mu = 0.05$, no time effect



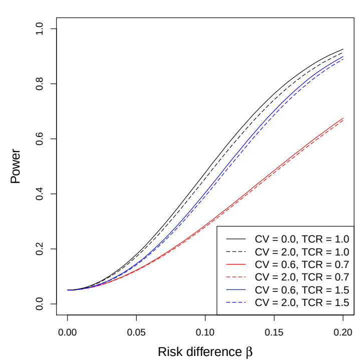
(b) $\mu = 0.6$, no time effect



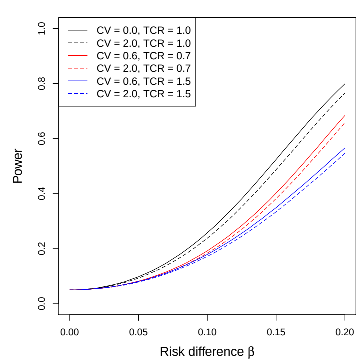
(c) $\mu = 0.05$, very small time effect



(d) $\mu = 0.6$, very small time effect

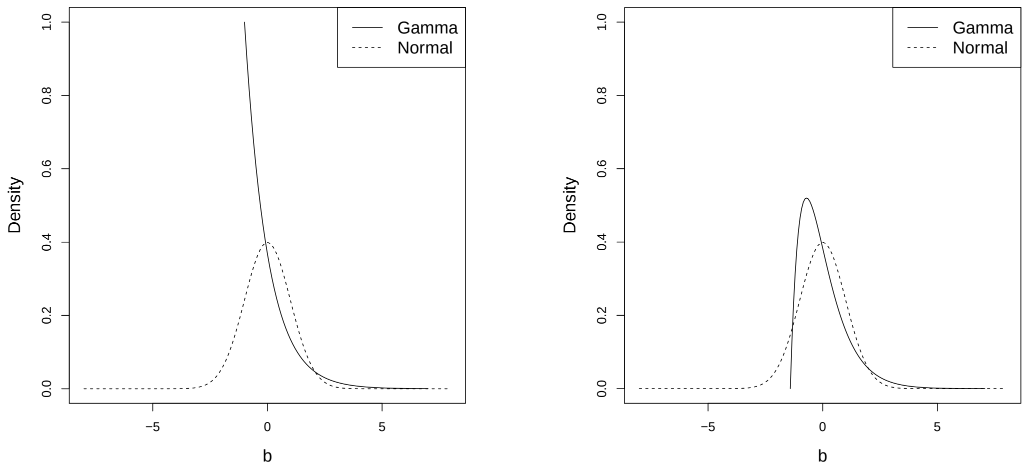


(e) $\mu = 0.05$, moderate time effect



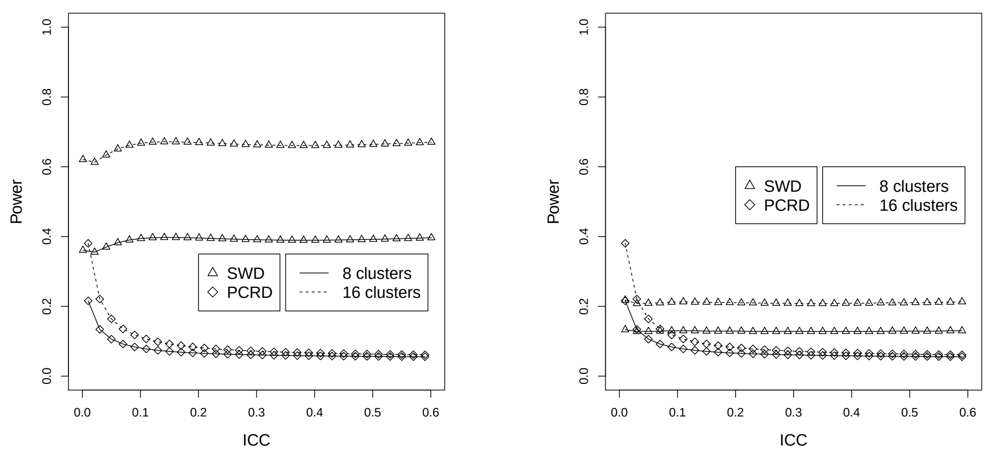
(f) $\mu = 0.6$, moderate time effect

Fig. S1. Plots of power v.s. the risk difference β, for different baseline risks and different values of CV and TCR, with the number of steps $J = 3$, mean cluster size $\bar{N} = 30$, number of clusters $I = 16$, and $ICC = 0.01$. There are no time effects ($\delta = 0$) in the first row, very small time effects ($\delta = 0.0001$) in the second row, and moderate time effects ($\delta = 0.05$) in the third row.



(a) Shape $\lambda = 1$ (b) Shape $\lambda = 2$

Fig. S2. Gamma and standard normal densities with the same mean and variance.



(a) power vs. ICC with $\mu = 0.2$, no time effect (b) power vs. ICC with $\mu = 0.2$, moderate time effect

Fig. S3. Comparison between SWD and pCRD. Power vs. ICC, for different baseline risks μ, where the risk difference $\beta = 0.05$, the cluster size $N = 90$, and the number of steps $J = 3$. (a) There are no time effects included in the model. (b) There are moderate time effects in the model.

Table S1. Finding the required number of partitions for the model with time effects in the PPIUD study in Tanzania ($N = 3600$), with different time effects and hypothetical risk ratios

	Time effects					
	$\delta = -0.0001$		$\delta = -0.0091$		$\delta = -0.0181$	
	Risk ratio		Risk ratio		Risk ratio	
	0.8	0.9	0.8	0.9	0.8	0.9
Q=16	0.205	0.068	0.233	0.109	0.305	0.252
Q=32	0.901	0.447	0.970	0.471	0.994	0.477
Q=64	0.973	0.471	0.982	0.498	0.983	0.504
Q=128	0.976	0.480	0.981	0.494	0.984	0.506

S3. The power of SWD and pCRD when there are no time effects

When ICC goes to zero, the variance of cluster random effect $\tau^2 \to 0$. Using the fact that $\lim_{\tau\to 0} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{b_i^2}{2\tau^2}} = \delta(b_i)$ and $\int_{-\infty}^{\infty} f(b_i)\delta(b_i)db_i = f(0)$, where $\delta(\cdot)$ is the Dirac delta function (Pathak, 1993), by the dominated convergence theorem, the log-likelihood (2.4) of the main paper goes to $\ell_0(\boldsymbol{\beta})$ as follows,

$$\ell_0(\boldsymbol{\theta}) = \sum_{i=1}^{I} \log \left(\prod_{n=1}^{N} (g^{-1}(\mu + \beta X_{in}))^{Y_{in}} (1 - g^{-1}(\mu + \beta X_{in}))^{1-Y_{in}} \right). \tag{S3.1}$$

When $g(\cdot)$ is an identity link, (S3.1) can be written as follows using $\mathbf{Z}_i = (Z_{i00}, Z_{i01}, Z_{i10}, Z_{i11})$,

$$\ell_0(\boldsymbol{\theta}) = \sum_{i=1}^{I} \log \left((1-\mu)^{Z_{i00}} (\mu)^{Z_{i01}} (1 - (\mu + \beta))^{Z_{i10}} (\mu + \beta)^{Z_{i11}} \right). \tag{S3.2}$$

The likelihood (S3.2) is the standard binomial likelihood for the independent two group design. We derive an expression for the MLE by setting its first derivative equal to zero. Then, the covariance matrix is given by the inverse of the expected value of the Hessian matrix of $\ell_0(\boldsymbol{\beta})$. Skipping the algebra here, it can be shown that the asymptotic variance of the MLE $Var(\hat{\beta}) = \frac{\mu(1-\mu)}{IN/2} + \frac{(\mu+\beta)(1-\mu-\beta)}{IN/2}$, which is exactly equal to (3.20) of the manuscript when the ICC goes to zero. Hence, the MLE from the SWD and the pCRD converge to the same estimator and so does the power of the two designs, as shown in Figures 4(e) of the manuscript and S3(a). It is well known that the pCRD is strongly sensitive to the ICC and that its power decreases as the ICC increases, while the SWD is relatively insensitive to the ICC (Hussey and Hughes, 2007), as shown in Figures 4(e) of the manuscript and S3(a). Thus, we have shown that the power of the SWD based on the MLE variance (2.11) of manuscript is always greater than that of the pCRD based on variance given by (3.20) of the main paper.

References

Hussey, Michael A and Hughes, James P. (2007, Feb). Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* **28**(2), 182–191.

Pathak, R. S. (1993). *Generalized functions and their applications*. New York: Plenum Press.