# Web-based Supplementary Materials for "Finite-Sample Corrected GEE of Population Average Treatment Effects in Stepped Wedge Cluster Randomized Trials" by JoAnna Scott, Allan deCamp, Michal Juraska, Michael P. Fay, Peter B. Gilbert

# 1 Web Appendix A: Standard GEE and Corrected GEE for a Small Numbers of Clusters

## 1.1 Standard GEE

Given some method for obtaining the estimates $Y_{ij}$, we consider a GEE approach to estimation in the glm $g(\mu_{ij}) = X_{ij}\beta$ with $g(\cdot)$ a known link function and $\beta$ a $p$-dimensional vector of regression coefficients, and $X_{ij} = (X_{i1}, \cdots, X_{iJ_i})^T$ a general $p$-dimensional covariate vector for cluster-step $i, j$ for $i = 1, \cdots, I$ and $j = 1, \cdots, J_i$ (generalized slightly from the main article which assumes $J_i = J$ for all $i$).

Let $\nu_{ij}$ be the conditional variance of $Y_{ij}$, $\nu_{ij} = V(Y_{ij}|X_{ij}) = \phi \cdot h(\mu_{ij})$, where $\phi$ is the dispersion parameter and $h(\cdot)$ is the variance function. Let $\Omega_i = V(Y_i|X_i)$ denote the conditional $J_i \times J_i$ covariance matrix. The GEE framework allows us to specify a working covariance matrix $\Sigma_i$ in place of the true covariance matrix $\Omega_i$. Given an estimate $\hat{\Sigma}_i$ of the working correlation matrix $\Sigma_i$, we estimate $\beta$ by solving the estimating equations

$$0 = \sum_{i=1}^{n} D_i^T \hat{\Sigma}_i^{-1}(Y_i - \mu_i),$$

where $\mu_i = (\mu_{i1}, \ldots, \mu_{iJ_i})^T$ and $D_i = \frac{\partial \mu_i}{\partial \beta^T}$. The covariance matrix $V(\hat{\beta})$ of $\hat{\beta}$ is consistently estimated by the sandwich estimator of variance,

$$V(\hat{\beta}) = A^{-1}BA^{-1} \tag{1}$$

with

$$A = \textstyle\sum_{i=1}^{n} D_i^T \Sigma_i^{-1} D_i \quad \text{and} \quad B = \sum_{i=1}^{n} D_i^T \Sigma_i^{-1} \Omega_i \Sigma_i^{-1} D_i \ .$$

When the working variance is correctly specified, the sandwich variance matrix of equation 1 simplifies to the Fisher information matrix $A^{-1}$. Whereas the Fisher information matrix is the bread of equation 1, the cheese portion, $B$, is called the outer product gradient (OPG). For estimation, residuals $Y_i - \mu_i$ are used to estimate $\Omega_i$, which is replaced in equation 1 by $\hat{\Omega}_i = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T$.

As shown in Mancl and DeRouen[1],

$$E[\hat{\Omega}] \approx (I_i - H_{ii})\Omega_i(I_i - H_{ii})^T + \sum_{i \neq j} H_{ij}\Omega_j H_{ij}^T, \tag{2}$$

where $H_{ij} = D_i A^{-1} D_j^T \Sigma_j^{-1}$ and $I_i$ is the $J_i \times J_i$ identity matrix.

## 1.2 Bias corrected sandwich variance GEE

We consider four different bias correction methods.

1. Based on the expectation shown in equation 2, and the assumption that the sum in that equation is negligible, Mancl and DeRouen[1] (henceforth MD) propose the following adjustment to the usual OPG estimator:

$$\hat{B}_{MD} = \sum_{i=1}^{n} \hat{D}_i^T \hat{\Sigma}_i^{-1}(I_i - \hat{H}_{ii})^{-1}\hat{\Omega}_i(I_i - \hat{H}_{ii})^{-1}\hat{\Sigma}_i^{-1}\hat{D}_i.$$

2. Second, Kauermann and Carroll[2] (KD) propose adjusting the OPG estimator using

$$\hat{B}_{KC} = \sum_{i=1}^{n} \hat{D}_i^T \hat{\Sigma}_i^{-1}(I_i - \hat{H}_{ii})^{-1/2}\hat{\Omega}_i(I_i - \hat{H}_{ii})^{-1/2}\hat{\Sigma}_i^{-1}\hat{D}_i.$$

   Kauermann and Carrol include the sum in equation 2 and assume that $\Sigma_i$ is correctly specified while Mancl and DeRouen do not, which accounts for the different derivations. As noted by Lu et al.[3], the elements of $H_{ii}$ are between 0 and 1, such that $\hat{B}_{KC}$ are expected to give larger standard errors than the usual OPG estimator, $\hat{B}_{OPG}$, and $\hat{B}_{MD}$ is larger than $\hat{B}_{KC}$.

   We show in Web Appendix B, that this KC bias adjustment is identical to the modified Fay and Graubard (mFG) bias correction of Ziegler[4], given by

$$\hat{B}_{mFG} = \sum_{i=1}^{n} \tilde{H}_i \hat{D}_i^T \hat{\Sigma}_i^{-1}\hat{\Omega}_i\hat{\Sigma}_i^{-1}\hat{D}_i\tilde{H}_i^T,$$

   where $\tilde{H}_i = (I - \hat{A}_i\hat{A}^{-1})^{-1/2}$, $I$ is the $p \times p$ identity matrix, and $\hat{A}_i = \hat{D}_i^T\hat{\Sigma}_i^{-1}\hat{D}_i$.

3. Third, the original bias correction of Fay and Graubard used

$$\hat{B}_{FG} = \sum_{i=1}^{n} \dot{H}_i \hat{D}_i^T \hat{\Sigma}_i^{-1}\hat{\Omega}_i\hat{\Sigma}_i^{-1}\hat{D}_i\dot{H}_i^T,$$

   where $\dot{H}_i$ is a diagonal matrix with $jj$th element equal to $\left\{1 - \min\left(0.75, \left[\hat{A}_i\hat{A}^{-1}\right]_{jj}\right)\right\}^{-1/2}$.

4. Fourth, the proposed correction of Morel et al.[5] (MBN) departs from the three earlier proposed corrections in that they use an additive correction term. Let $\hat{V}(\hat{\beta})$ be the usual plug in sandwich estimator using $\hat{B}_{OPG}$. Then

$$\hat{V}_{MBN}(\hat{\beta}) = \hat{V}(\hat{\beta}) + \hat{\delta}_n \cdot \hat{\zeta} \cdot \hat{A}^{-1},$$

   where $\hat{\delta}_n = \min\left(\frac{1}{2}, \frac{p}{n-p}\right)$ and $\hat{\zeta} = \max(1, \mathrm{trace}(\hat{A}^{-1}\hat{B}_{OPG})/p)$.

As shown in Web Appendix B, the KC and mFG variance-covariance matrix estimators are identical. However, their calculation involves numerical inversion of different matrices. For the KC variance estimator this matrix has dimension $J_i \times J_i$ whereas for the mFG estimator it is dimension $p \times p$. Numerical algorithms for matrix inversion can give slightly different answers, and in practice the choice of implementation might be driven by the size of the typical cluster versus the number of covariates in the model, where a small cluster size would likely result in more stability in the numeric computation.

### 1.3 The Fay and Graubard $\delta_5$ (FG d5) method

The Fay and Graubard $\delta_5$ (FG d5) method, uses the third bias correction of the previous section, together with using a different t-distribution for each parameter. The degrees of freedom for the $j$th parameter is

$$\tilde{d}_j = \frac{\left\{\text{trace}\left(\tilde{\Psi}_j G^T M_j G\right)\right\}^2}{\text{trace}\left(\tilde{\Psi}_j G^T M_j G \tilde{\Psi}_j G^T M_j G\right)},$$

where

$$\begin{aligned}
\tilde{\Psi}_j &= \text{ block diagonal, with } k\text{th block equal } \tilde{\Psi}_{jk} \\
\tilde{\Psi}_{jk} &= \frac{w_{jk}\sum_{i=1}^n \hat{\Psi}_i}{\sum_{h=1}^n w_{jh}} \\
w_{jk} &= C_j^T \left\{ \left( \sum_{i \neq k} \hat{A}_i \right)^{-1} - \hat{A}^{-1} \right\} C_j \\
C_j &= \text{ a } p \times 1 \text{ vector of zeros except the } j\text{th element is equal to 1} \\
\hat{\Psi}_i &= \dot{H}_i \left( \hat{D}_i^T \hat{\Sigma}_i^{-1} (Y_i - \hat{\mu}_i) \right) \left( \hat{D}_i^T \hat{\Sigma}_i^{-1} (Y_i - \hat{\mu}_i) \right)^T \dot{H}_i^T \\
G &= I_{np} - \begin{bmatrix} \hat{A}_1 \\ \cdots \\ \hat{A}_n \end{bmatrix} \hat{A}^{-1} [I_p, \ldots, I_p] \\
M_j &= \text{ block diagnonal matix with } i\text{th block } \dot{H}_i \hat{A}^{-1} C_j C_j^T \hat{A}^{-1} \dot{H}_i^T.
\end{aligned}$$

## 2 Web Appendix B: KC and mFG Adjusted Variance-Covariance Matrix Estimators are Identical

To show the desired result, we first simplify the notation for the OPG adjustments used in the two procedures. Throughout this section $I$ is an identity matrix with dimension implied by the context in which it is used. We seek to show that the KC adjustment factor

$$\hat{D}_i^T \hat{\Sigma}_i^{-1} (I - \hat{H}_{ii})^{-1/2} = \hat{D}_i^T \hat{\Sigma}_i^{-1} (I - \hat{D}_i \hat{A}^{-1} \hat{D}_i^T \hat{\Sigma}_i^{-1})^{-1/2}$$

is equal to the mFG adjustment factor

$$\tilde{H}_i \hat{D}_i^T \hat{\Sigma}_i^{-1} = (I - \hat{D}_i^T \hat{\Sigma}_i^{-1} \hat{D}_i \hat{A}^{-1})^{-1/2} \hat{D}_i^T \hat{\Sigma}_i^{-1}.$$

To simplify notation we define

$$\begin{aligned}
X &= \hat{D}_i^T \hat{\Sigma}_i^{-1} \\
Y &= \hat{D}_i \\
Z &= \hat{A}^{-1} \\
W &= (I - YZX)^{-1} \\
V &= (I - XYZ)^{-1}.
\end{aligned}$$

In simplified notation we would like to show that $XW^{1/2} = V^{1/2}X$.

First we show that $XW = VX$ using the Sherman-Morrison-Woodbury identity.

$$\begin{aligned}
XW = X(I - YZX)^{-1} &= X\left[I + Y(Z^{-1} - XY)^{-1}X\right] \\
&= \left[I + XY(Z^{-1} - XY)^{-1}\right] X \\
&= (I - XYZ)^{-1}X \\
&= VX
\end{aligned}$$

Then, by induction we can show that

$$X(W - I)^n = (V - I)^n X$$

for all integers $n$. The identity above proves the case for $n = 1$, and given $X(W - I)^k = (V - I)^k X$,

$$\begin{aligned}
X(W - I)^{k+1} &= X(W - I)^k (W - I) \\
&= (V - I)^k X(W - I) \\
&= (V - I)^k (V - I)X \\
&= (V - I)^{k+1} X,
\end{aligned}$$

which proves the induction step.

Finally, using a Taylor expansion of $W^{1/2}$ we have

$$\begin{aligned}
XW^{1/2} &= X\left[I + \frac{1}{2}(W - I) + \sum_{n=2}^{\infty}(-1)^{n+1}\frac{(2n-3)!}{2^{n-2}n!(n-2)!}(W - I)^n\right] \\
&= \left[I + \frac{1}{2}(V - I) + \sum_{n=2}^{\infty}(-1)^{n+1}\frac{(2n-3)!}{2^{n-2}n!(n-2)!}(V - I)^n\right] X \\
&= V^{1/2}X.
\end{aligned}$$

# 3 Web Appendix C: Survey of Numbers of Clusters used in Published Stepped Wedge Cluster Randomized Trials

A literature search was conducted to identify all publications describing stepped wedge cluster randomized trials, in order to assess the distribution of the number of randomized clusters used in these trials. Restricting to all published stepped wedge CRTs with at least 50 individuals per cluster on average showed that the mean and median number of clusters was 33 and 12, respectively, with interquartile range 7–29.

Table 1: Number of clusters used in Stepped Wedge Cluster Randomized Trials.

| | Publication Reference | Total Number of Subjects | Number of Clusters | Average Number of Subjects per Cluster |
|---|---|---|---|---|
| 1. | Haugen et al. [2014][6] | 5295 | 5 | 1059.00 |
| 2. | Franklin et al. [2014][7] | 16357 | 15 | 1090.47 |
| 3. | Bailey et al. [2014][8] | 6066 | 6 | 1011.00 |
| 4. | Fink et al. [2013][9] | 59905 | 33 | 1815.30 |
| 5. | Poldervaart et al. [2013][10] | 6600 | 10 | 660.00 |
| 6. | Chinbuah et al. [2013][11] | 12333 | 114 | 108.18 |
| 7. | Durovni et al. [2013][12] | 12816 | 29 | 441.93 |
| 8. | Bashour et al. [2013][13] | 2000 | 4 | 500.00 |
| 9. | Doherty et al. [2013][14] | 16627 | 45 | 369.49 |
| 10. | Marshall et al. [2012][15] | 6250 | 26 | 240.38 |
| 11. | Gucciardi et al. [2012][16] | 1200 | 12 | 100.00 |
| 12. | Solomon et al. [2012][17] | 6400 | 128 | 50.00 |
| 13. | van den Broek et al. [2012][18] | 43358 | 190 | 228.20 |
| 14. | Roy et al. [2013][19] | 1315 | 24 | 54.79 |
| 15. | Mouchoux et al. [2011][20] | 360 | 3 | 120.00 |
| 16. | Lilly et al. [2011][21] | 6290 | 7 | 898.57 |
| 17. | Turner et al. [2011][22] | 600 | 5 | 120.00 |
| 18. | Killam et al. [2010][23] | 1566 | 8 | 195.75 |
| 19. | Winani et al. [2007][24] | 3262 | 10 | 326.20 |
| 20. | Ciliberto et al. [2005][25] | 1178 | 7 | 168.29 |
| 21. | The Gambia Hepatitis Study Group [1987][26] | 120000 | 17 | 7058.82 |

# 4 Web Appendix D: Additional Simulation Results

All of the simulation studies reported in the main article and in this Web Appendix are based on 5000 simulated data sets. This number of iterations implies that the Monte Carlo standard error for the size and coverage probability is $\sqrt{0.05 * 0.95/5000} = 0.00308$. The estimated sizes and coverage probabilities are compared to an interval from -2*0.00308 and 2*0.00308 to help judge accuracy of the results.

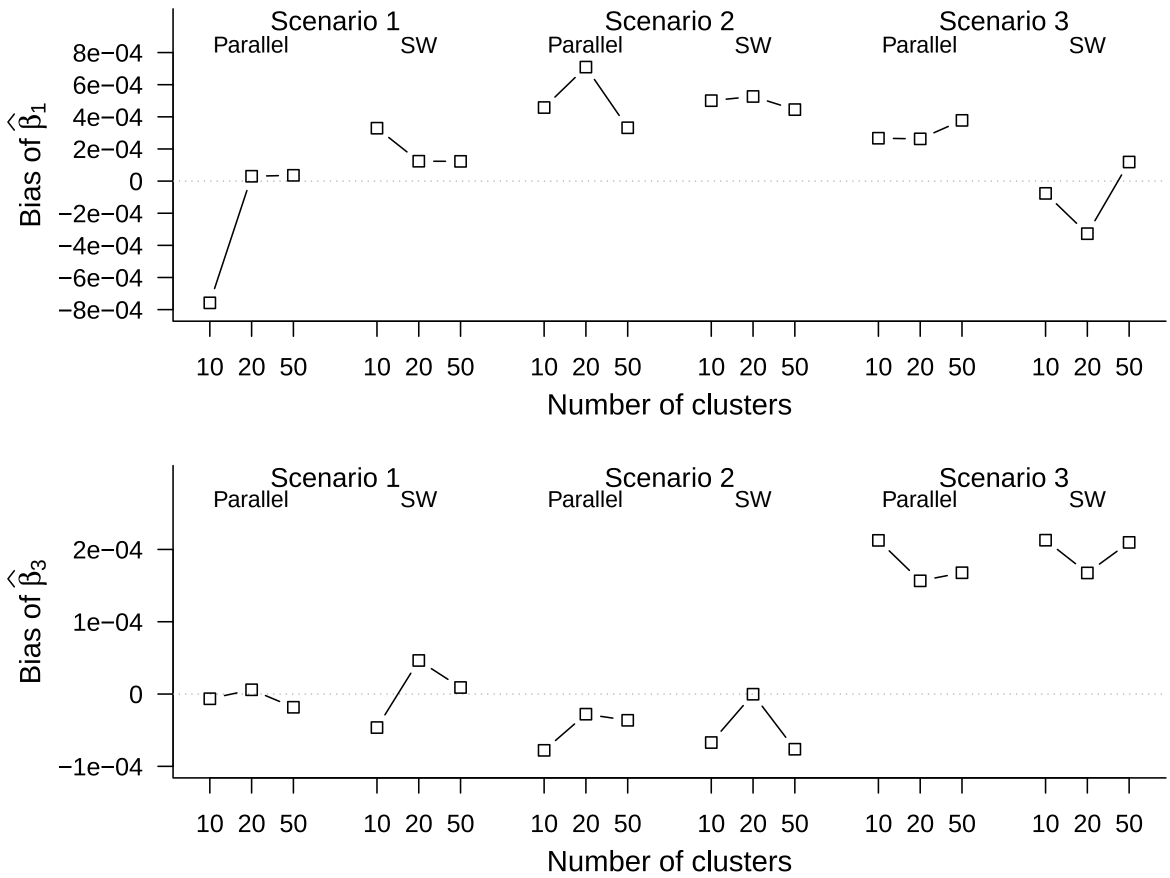


Figure 1: Finite-sample bias of estimation of $\beta_1$ and $\beta_3$ using the stepped wedge and parallel designs in simulation scenarios 1–3 with cluster-step incidences satisfying the AR-1 correlation structure. In scenarios 1–2, a reduced form of model (2) in the main article excluding the interaction term is considered for inference about $\beta_1$.

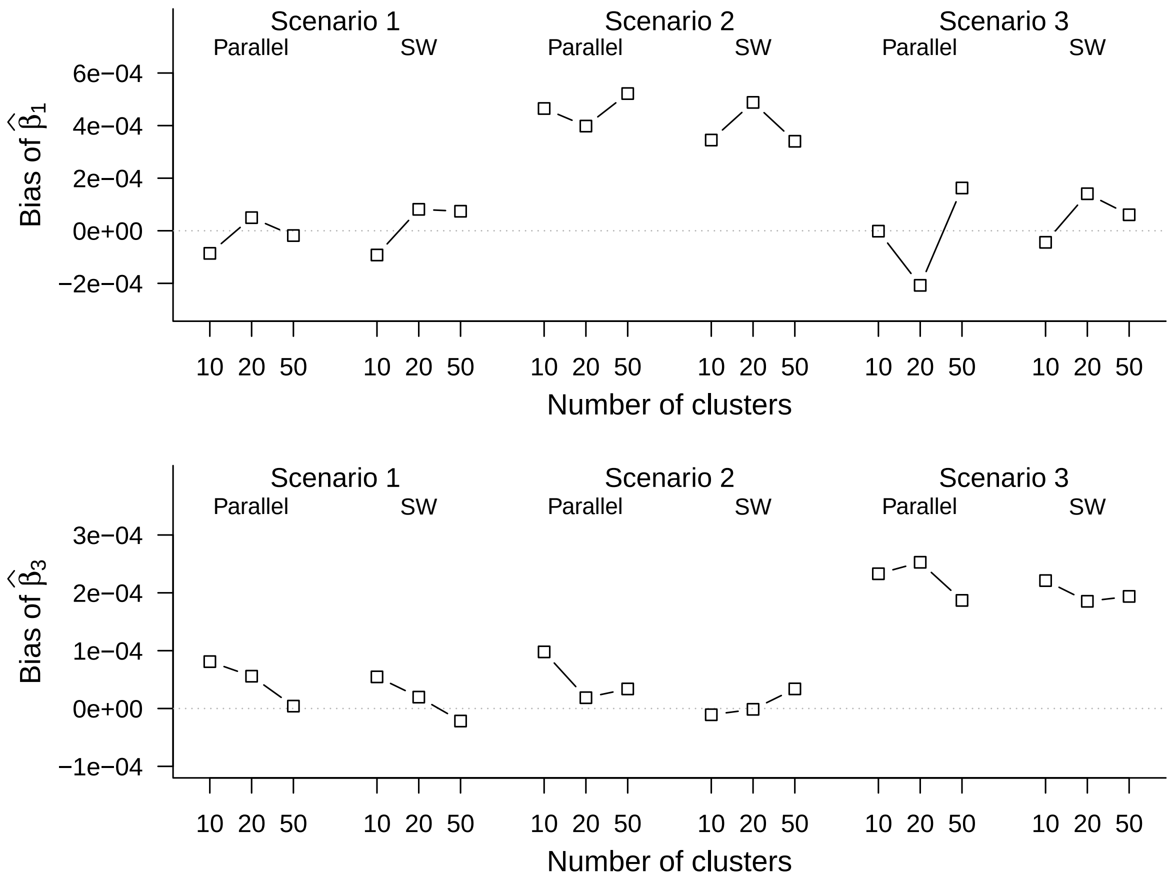


Figure 2: Finite-sample bias of estimation of $\beta_1$ and $\beta_3$ using the stepped wedge and parallel designs in simulation scenarios 1–3 with independent cluster-step incidences. In scenarios 1–2, a reduced form of model (2) in the main article excluding the interaction term is considered for inference about $\beta_1$.

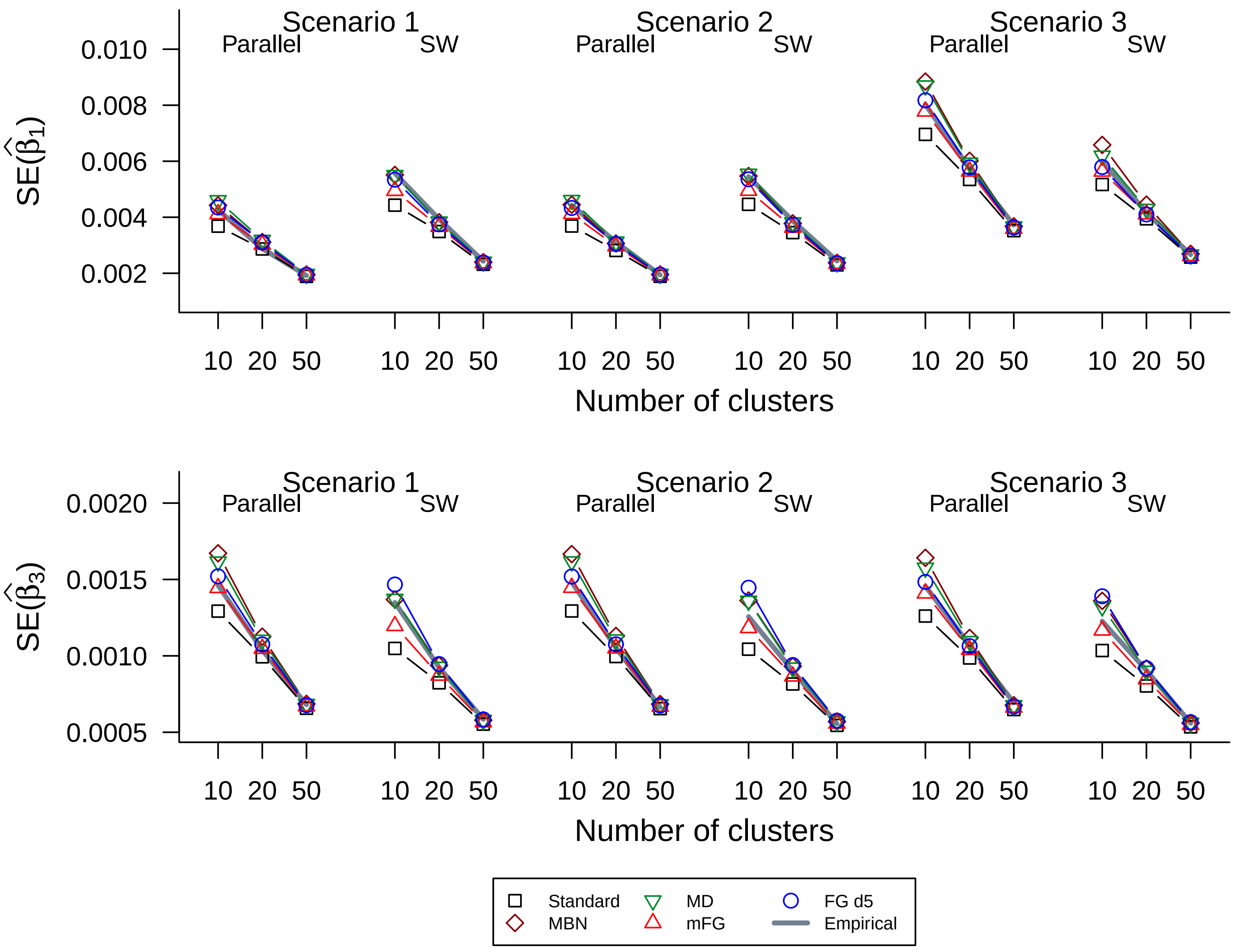


Figure 3: Median standard-GEE, bias-corrected and empirical standard error (SE) estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_3$ using the stepped wedge and parallel designs in simulation scenarios 1–3 with independent cluster-step incidences. The empirical SE estimate is computed as the sample standard deviation of the $\widehat{\beta}$ estimates. In scenarios 1–2, a reduced form of model (2) in the main article excluding the interaction term is considered for inference about $\beta_1$.

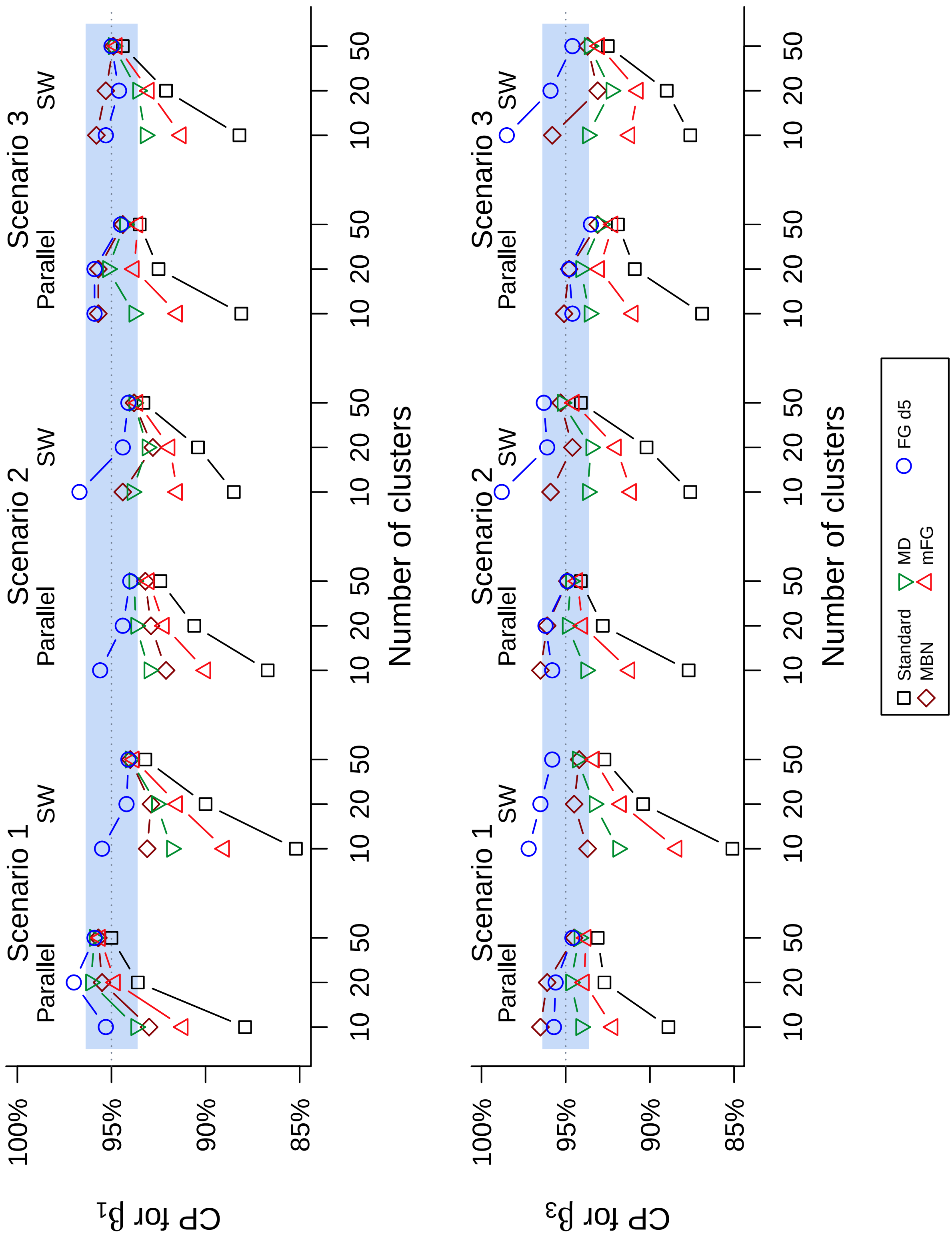


Figure 4: Coverage probabilities (CP) of 95% standard-GEE and bias-corrected Wald confidence intervals for $\beta_1$ and $\beta_3$ using the stepped wedge and parallel designs in simulation scenarios 1–3 with independent cluster-step incidences. The horizontal band represents $\pm 2 \times$ Monte Carlo standard error. In scenarios 1–2, a reduced model excluding the interaction term is considered for inference about $\beta_1$.

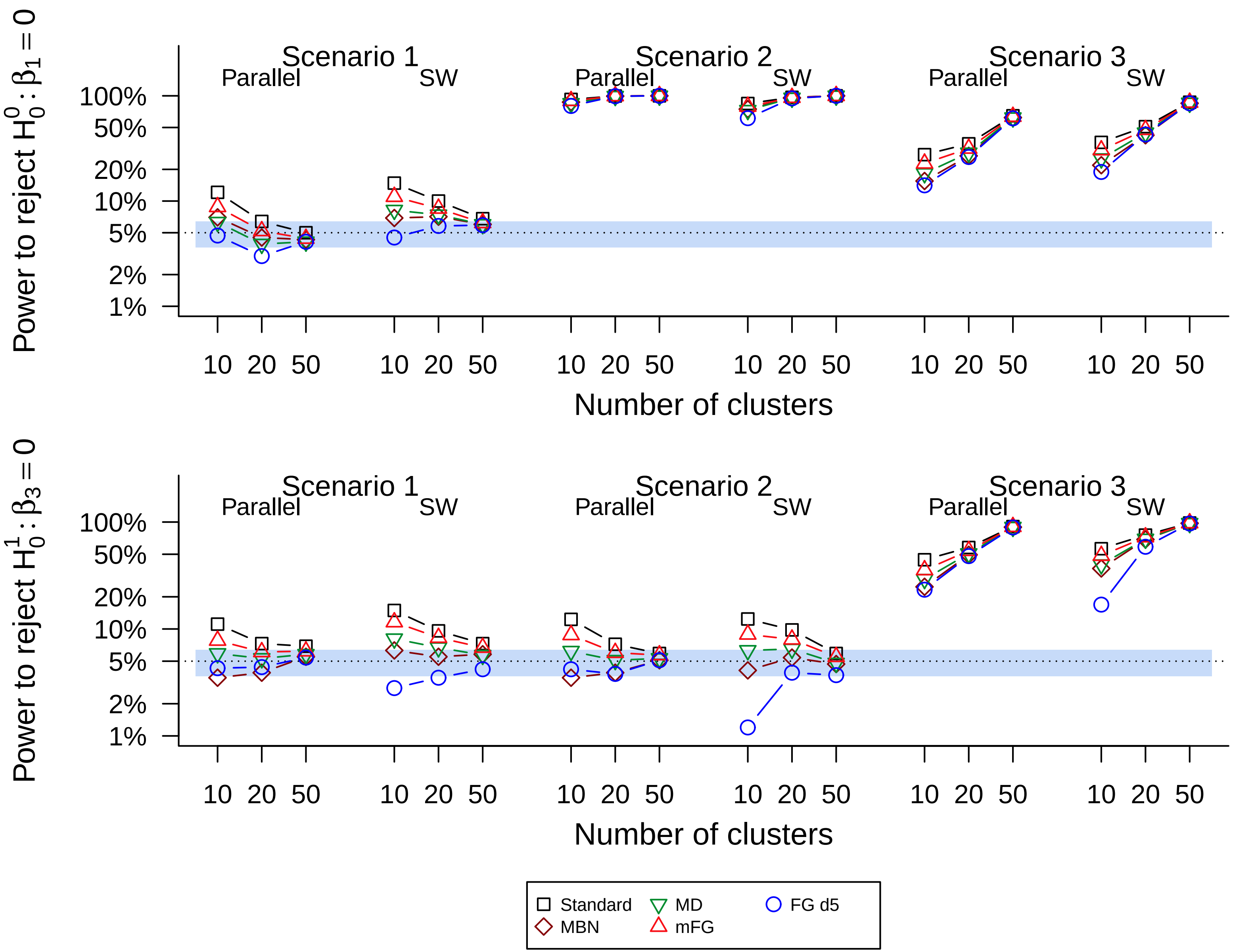


Figure 5: Size and power of standard-GEE and bias-corrected Wald tests to reject the null hypotheses $H_0^0 : \beta_1 = 0$ and $H_0^1 : \beta_3 = 0$ at 5% significance level using the stepped wedge and parallel designs in simulation scenarios 1–3 with independent cluster-step incidences. For readability of the estimated sizes, size and power are plotted on the log scale.

# References


[1] Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. Biometrics. 2001 March;57(1):126–134.

[2] Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. Journal of the American Statistical Association. 2001 December;96(456):1387–1396.

[3] Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala SI, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. Biometrics. 2007 September;63(3):935–941.

[4] Ziegler A. Generalized Estimating Equations. New York: Springer; 2011.

[5] Morel JG, Bokossa MC, Neerchal NK. Small sample correction for the variance of GEE estimators. Biometrical Journal. 2003;45(4):395–409.

[6] Haugen AS, Softeland E, Almeland SK, Sevdalis N, Vonen B, Eide GE, et al. Effect of the World Health Organization checklist on patient outcomes: A stepped wedge cluster randomized controlled trial. Annals of Surgery. 2014 May 13;.

[7] Franklin BD, Reynolds M, Sadler S, Hibberd R, Avery AJ, Armstrong SJ, et al. The effect of the electronic transmission of prescriptions on dispensing errors and prescription enhancements made in English community pharmacies: A naturalistic stepped wedge study. BMJ Quality and Safety. 2014 April 17;.

[8] Bailey FA, Williams BR, Woodby LL, Goode PS, Redden DT, Houston TK, et al. Intervention to improve care at life's end in inpatient settings: The BEACON Trial. Journal of General Internal Medicine. 2014 June;29(6):836–843.

[9] Fink G, Robyn PJ, Sie A, Sauerborn R. Does health insurance improve health?: Evidence from a randomized community-based insurance rollout in rural Burkina Faso. Journal of Health Economics. 2013 Dec;32(6):1043–1056.

[10] Poldervaart JM, Reitsma JB, Koffijberg H, Backus BE, Six AJ, Doevendans PA, et al. The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial. BMC Cardiovasc Disord. 2013;13:77.

[11] Chinbuah MA, Adjuik M, Cobelens F, Koram KA, Abbey M, Gyapong M, et al. Impact of treating young children with antimalarials with or without antibiotics on morbidity: a cluster-randomized controlled trial in Ghana. Int Health. 2013 Sep;5(3):228–235.

[12] Durovni B, Saraceni V, Moulton LH, Pacheco AG, Cavalcante SC, King BS, et al. Effect of improved tuberculosis screening and isoniazid preventive therapy on incidence of tuberculosis and death in patients with HIV in clinics in Rio de Janeiro, Brazil: a stepped wedge, cluster-randomised trial. Lancet Infect Dis. 2013 Oct;13(10):852–858.

[13] Bashour HN, Kanaan M, Kharouf MH, Abdulsalam AA, Tabbaa MA, Cheikha SA. The effect of training doctors in communication skills on women's satisfaction with doctor-woman relationship during labour and delivery: a stepped wedge cluster randomised trial in Damascus. BMJ Open. 2013;3(8).

[14] Doherty S, Knott J, Bennetts S, Jazayeri M, Huckson S. National project seeking to improve pain management in the emergency department setting: findings from the NHMRC-NICS National Pain Management Initiative. Emerging Medicine Australas. 2013 Apr;25(2):120–126.

[15] Marshall T, Caley M, Hemming K, Gill P, Gale N, Jolly K. Mixed methods evaluation of targeted case finding for cardiovascular disease prevention using a stepped wedged cluster RCT. BMC Public Health. 2012;12:908.

[16] Gucciardi E, Fortugno M, Horodezny S, Lou W, Sidani S, Espin S, et al. Will Mobile Diabetes Education Teams (MDETs) in primary care improve patient care processes and health outcomes? Study protocol for a randomized controlled trial. Trials. 2012;13:165.

[17] Solomon E, Rees T, Ukoumunne OC, Hillsdon M. The Devon Active Villages Evaluation (DAVE) trial: study protocol of a stepped wedge cluster randomised trial of a community-level physical activity intervention in rural southwest England. BMC Public Health. 2012;12:581.

[18] van den Broek IV, van Bergen JE, Brouwers EE, Fennema JS, Gotz HM, Hoebe CJ, et al. Effectiveness of yearly, register based screening for chlamydia in the Netherlands: controlled trial with randomised stepped wedge implementation. BMJ. 2012;345:e4316.

[19] Roy A, Anaraki S, Hardelid P, Catchpole M, Rodrigues LC, Lipman M, et al. Universal HIV testing in London tuberculosis clinics: a cluster randomised controlled trial. European Respiratory Journal. 2013 Mar;41(3):627–634.

[20] Mouchoux C, Rippert P, Duclos A, Fassier T, Bonnefoy M, Comte B, et al. Impact of a multifaceted program to prevent postoperative delirium in the elderly: the CONFUCIUS stepped wedge protocol. BMC Geriatr. 2011;11:25.

[21] Lilly CM, Cody S, Zhao H, Landry K, Baker SP, McIlwaine J, et al. Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. JAMA. 2011 Jun;305(21):2175–2183.

[22] Turner J, Kelly B, Clarke D, Yates P, Aranda S, Jolley D, et al. A randomised trial of a psychosocial intervention for cancer patients integrated into routine care: the PROMPT study (promoting optimal outcomes in mood through tailored psychosocial therapies). BMC Cancer. 2011;11:48.

[23] Killam WP, Tambatamba BC, Chintu N, Rouse D, Stringer E, Bweupe M, et al. Antiretroviral therapy in antenatal care to increase treatment initiation in HIV-infected pregnant women: a stepped-wedge evaluation. AIDS. 2010 Jan;24(1):85–91.

[24] Winani S, Wood S, Coffey P, Chirwa T, Mosha F, Changalucha J. Use of a clean delivery kit and factors associated with cord infection and puerperal sepsis in Mwanza, Tanzania. J Midwifery Womens Health. 2007;52(1):37–43.

[25] Ciliberto MA, Sandige H, Ndekha MJ, Ashorn P, Briend A, Ciliberto HM, et al. Comparison of home-based therapy with ready-to-use therapeutic food with standard therapy in the treatment of malnourished Malawian children: a controlled, clinical effectiveness trial. American Journal of Clinical Nutrition. 2005;81:864–870.

[26] The Gambia Hepatitis Study Group. The Gambia Hepatitis Intervention Study. Cancer Research. 1987 November 1;47:5782–5787.