Power Calculations for Stepped Wedge Designs with Binary Outcomes: Methods and Software

Presented by:
Xin Zhou, Ph.D.
Yale School of Public Health

If you’re a person with a disability or using assistive technology and are having difficulty accessing any of the content in this file, please contact the ODP at prevention@nih.gov.
Power Calculations for Stepped Wedge Designs with Binary Outcomes: Methods and Software

Xin Zhou

Center for Methods in Implementation and Prevention Science (CMIPS)
Department of Biostatistics

Yale School of Public Health
Promotes the adoption of research findings into clinical practice and develops and assesses sustainable interventions to improve public health

Serves as an international leader in methods innovation for public health, providing biostatistical and methodological expertise on new implementation science projects

**CMIPS Biostatistics Team**

**Donna Spiegelman, ScD**, Director and Susan Dwight Bliss Professor of Biostatistics

**Xin Zhou, PhD**, Assistant Professor of Biostatistics

**Fan Li, PhD**, Assistant Professor of Biostatistics

**Laura Forastiere, PhD**, Assistant Professor of Biostatistics

Additional CMIPS primary and associate faculty specialize in qualitative research and ethnography, health economics, maternal and child health promotion, global mental health promotion, cervical cancer prevention, and other areas of priority for the Center.
Outline

- Introduction of stepped wedge designs (SWDs)
- Existing power calculation methods for SWDs with binary outcomes
- Power calculation using a maximum likelihood approach for SWDs with binary outcomes
- Power calculation using generalized estimating equations for SWDs with binary outcomes
- Software
During the year following the birth of a child, 40% of women are estimated to have an unmet need for contraception.

Postpartum intrauterine device (PPIUD) meets women’s need for contraceptive protection following childbirth

- Long-term contraceptive protection
- Reversible
- Safe, effective and convenient
- Does not interfere with breastfeeding during the postpartum period

Uptake of PPIUD is low

- PPIUD insertion should be performed by a trained provider in the early postpartum period
- The services are not widely available

The International Federation of Gynaecology and Obstetrics (FIGO) implements an intervention that institutionalize PPIUD services

- as a routine part of antenatal counselling
- as a part of delivery room services

---

Cluster randomized trials

- Intervention requires application at a cluster level
- Avoid contamination (e.g. sharing medication or health messages)
Cluster randomized trials

- Different types of cluster randomized trials (CRT)

- Cross-sectional v.s. cohort designs
**Stepped wedge design**

- **Stepped Wedge Design (SWD)**

![Diagram showing stepped wedge design with clusters and intervention/control periods.](image)

- **Ethics:** all clusters need to get the potentially efficacious treatment
- **Feasibility:** not able to give the intervention simultaneously to many clusters

---

Motivating example

Postpartum intrauterine device (PPIUD) study design

- Collaboration with
  - International Federation of Gynaecology and Obstetrics (FIGO)
  - Association of Gynaecologists and Obstetricians of Tanzania (AGOTA)
  - Harvard T.H. Chan School of Public Health (HSPH)

- Impact of PPIUD use in Tanzania
Motivating example

Postpartum intrauterine device (PPIUD) study design

- Power to detect the effectiveness of PPIUD in preventing unintended pregnancy for 1.5 years following the index birth
- Enrollment is one year long.
- There are 6 clusters (hospitals), and 4 time periods.
- Each time period is 3 months long for enrollment.
- Approximately 900 women join each hospital in three months
- The cluster size is expected to be $900 \times 4 = 3,600$.

![Diagram showing stepped wedge design with binary outcomes](image)
Statistical power is an important component of study design

The **power** of a statistical test $H_0$ v.s. $H_A$:

\[
\text{Power} = \Pr(\text{reject } H_0 \mid H_A \text{ is true})
\]

<table>
<thead>
<tr>
<th>Type I error (false positive): the error of rejecting $H_0$ when it is actually true.</th>
<th>$\alpha = \Pr(\text{Type I error})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type II error (false negative): the error of failing to reject $H_0$ when $H_A$ is true.</td>
<td>$\beta = \Pr(\text{Type II error})$</td>
</tr>
</tbody>
</table>

\[
\text{Power} = 1 - \beta
\]
Existing power calculation method

Hussey and Hughes (2007)

- Apply a linear mixed model (LMM) for continuous outcomes

\[ Y_{ijk} = \mu + X_{ij} \beta + \gamma_j + b_i + \epsilon_{ijk}, \]

where \( Y_{ijk} \) is the outcome of individual \( k(\in \{1, \cdots, K\}) \) at time period \( j(\in \{1, \cdots, J\}) \) from cluster \( i(\in \{1, \cdots, I\}) \),

- \( X_{ij} \) is an indicator of the treatment (1=intervention; 0=control),
- \( \mu \) is the expected outcome at baseline (control at period 1), \( \beta \) is the treatment effect,
- \( \gamma_j \) is the time effect corresponding to period \( j \) (\( \gamma_1 = 0 \) for identifiability),
- \( \epsilon_{ijk} \sim N(0, \sigma_e^2) \) is the error term, \( b_i \sim N(0, \tau^2) \) is the random effect for cluster \( i \)


Xin Zhou (Yale Biostatistics) Stepped Wedge Designs with Binary Outcomes 10 / 35
Existing power calculation method

Hussey and Hughes (2007)

- Power calculation is based on a Wald test for
  \[ H_0 : \beta = 0 \text{ versus } H_A : \beta = \beta_A. \]

- The power for conducting a two-tailed test of size \( \alpha \) is
  \[
  \text{power} = \Phi \left( \frac{|\beta_A|}{\sqrt{\text{Var}(\hat{\beta})}} - Z_{1-\alpha/2} \right),
  \]
  where \( \Phi(\cdot) \) is the cumulative standard Normal distribution,
  \( Z_{1-\alpha/2} \) is the \((1 - \alpha/2)'th\) quantile of the standard Normal distribution.

- The estimated treatment effect \( \hat{\beta} \) is obtained from a weighted least squares (WLS)
  analysis, assuming that \( \sigma^2_e \) and \( \tau^2 \) are known.
Hussey and Hughes (2007)

- The linear mixed model (LMM) for continuous outcomes

\[ Y_{ijk} = \mu + X_{ij}\beta + \gamma_j + b_i + \epsilon_{ijk}, \]

where \( \epsilon_{ijk} \sim N(0, \sigma_e^2) \) is the error term, \( b_i \sim N(0, \tau^2) \) is the random effect for cluster \( i \).

- The variance of \( \hat{\beta} \) is \(^1\)

\[ \text{Var}(\hat{\beta}) = \frac{I\sigma^2(\sigma^2 + J\tau^2)}{(IU - W)\sigma^2 + (U^2 + IJU - JW - IV)\tau^2}, \]

where \( \sigma^2 = \sigma_e^2 / K \), \( U = \sum_{ij} X_{ij} \), \( W = \sum_j (\sum_i X_{ij})^2 \), and \( V = \sum_i (\sum_j X_{ij})^2 \).

---

Existing power calculation method

Hussey and Hughes (2007)

- **Pros**
  - closed-form formula for power calculation
  - Simple and fast

- **Cons**
  - Not designed for binary outcomes

- **SWD power calculation software for binary outcomes**
  - Excel spreadsheet
    - For binary outcomes: [http://faculty.washington.edu/jphughes/Power%20for%20stepped%20wedge.xls](http://faculty.washington.edu/jphughes/Power%20for%20stepped%20wedge.xls)

  - Normal approximation: $\sigma_e^2 = [0.5(p_1 + p_2)][1 − 0.5(p_1 + p_2)]$.
  - PASS Sample Size Software: several options on the approximation of the variance
    - $\sigma_e^2 = p_1(1 − p_1)$
    - $\sigma_e^2 = [0.5(p_1 + p_2)][1 − 0.5(p_1 + p_2)]$
    - $\sigma_e^2 = 0.5p_1(1 − p_1) + 0.5p_2(1 − p_2)$
Existing power calculation method

Hussey and Hughes’ approximation method for binary outcomes

- It is inaccurate.\(^1\)

- A toy example
  24 Clusters
  3 time periods
  Baseline proportion:
  \( p_1 = 50\% \)
  Proportion after intervention: \( p_2 = 40\% \)
  ICC \( \rho = 0.1 \)

- To achieve 80\% power
  H&H: required sample size 3,456
  MLE: required sample size 2,880

Method – a Maximum Likelihood Approach

- We consider a binary treatment $X_{ijk}$ and a binary outcome $Y_{ijk}$ for individual $k$ in cluster $i$ at time period $j$, where $i \in \{1, \cdots, I\}$, $j \in \{1, \cdots, J\}$, and $k \in \{1, \cdots, K\}$.

- The cluster size is $N = J \cdot K$, and the sample size is $I \cdot J \cdot K$.

- Two cases: without time effects and with time effects.

Method – a Maximum Likelihood Approach

The model with time effects

- Apply a generalized linear mixed model (GLMM) for binary outcomes

\[ g(p_{ijk}) = \mu + X_{ijk}\beta + \gamma_j + b_i, \]

where \( \mu \) is the probability of the outcome at baseline (control at period 1),
\( \beta \) is the treatment effect
\( \gamma_j \) is the time effect corresponding to period \( j \) (\( \gamma_1 = 0 \) for identifiability)
\( b_i \sim N(0, \tau^2) \) is the random effect for cluster \( i \),

\[ p_{ijk} = E(Y_{ijk}|X_{ijk}, b_i) = Pr(Y_{ijk} = 1|X_{ijk}, b_i) \]

- The parameters \( \theta = (\mu, \beta, \gamma_2, \cdots, \gamma_J, \tau^2) \).

- Link function \( g(\cdot) \)
  - Identity link – parameter of interest is the risk difference
  - Log link – parameter of interest is the risk ratio (relative risk)
  - Logit link – parameter of interest is odds ratio
The model without time effects

- Apply a generalized linear mixed model (GLMM) for binary outcomes

\[ g(p_{ijk}) = \mu + X_{ijk}\beta + b_i, \]

where \( \mu \) is the probability of the outcome at baseline (control at period 1),
\( \beta \) is the treatment effect,
\( b_i \sim N(0, \tau^2) \) is the random effect for cluster \( i \),
\( p_{ijk} = E(Y_{ijk}|X_{ijk}, b_i) = Pr(Y_{ijk} = 1|X_{ijk}, b_i) \)

- The parameters \( \theta = (\mu, \beta, \tau^2) \).
Method – a Maximum Likelihood Approach

Power calculation

- The asymptotic power for $H_0 : \beta = 0$ versus $H_A : \beta = \beta_A$

\[
power = \Phi \left( \frac{|\beta_A|}{\sqrt{\text{Var} (\hat{\beta})}} - Z_{1 - \alpha/2} \right).
\]

- In the likelihood theory, $\text{Var} (\hat{\beta})$ can be obtained by the inverse of information matrix.
Method – Generalized Estimating Equations

GEE model with time effects

- Apply a generalized estimating equations model (GEE) for binary outcomes

\[ g(p_{ijk}) = \mu + X_{ijk}\beta + \gamma_j, \]

where \( \gamma_j \) is the time effect corresponding to time period \( j \) (\( \gamma_1 = 0 \) for identifiability)

- \( p_{ijk} = P(Y_{ijk} = 1|X_{ijk}) \).

- The link function \( g(\cdot) \) can choose from: identity, log and logit links.

- The parameters \( \theta = (\mu, \beta, \gamma_2, \cdots, \gamma_J) \).

---

A correlation structure $R_i$ with three levels of correlation is employed:

- $\alpha_0$, within-period correlation ($\text{corr}(Y_{ijk}, Y_{ijk'}) = \alpha_0$, $k \neq k'$)
- $\alpha_1$, inter-period correlation ($\text{corr}(Y_{ijk}, Y_{ij'k'}) = \alpha_1$, $j \neq j'$)
- $\alpha_2$, within-individual correlation ($\text{corr}(Y_{ijk}, Y_{ij'k}) = \alpha_2$, $j \neq j'$)

(a) cross-sectional  

(b) closed cohort
Method – Generalized Estimating Equations

GEE model with time effects

- For cluster $i$, $Y_i = (Y_{i11}, Y_{i12}, \ldots, Y_{iJK})^T$ and $p_i = (p_{i11}, p_{i12}, \ldots, p_{iJK})^T$.

- The GEE estimator $\hat{\theta}$ is obtained by solving the equation

$$\sum_{i=1}^l D_i^T V_i^{-1} (Y_i - p_i) = 0.$$ 

$D_i = \partial p_i / \partial \theta$, $V_i = A_i^{1/2} R_i A_i^{1/2}$

- $A_i$ is the $N$-dimensional diagonal matrix with elements $v(p_{ijk})$

- The variance function is $v(p_{ijk}) = p_{ijk}(1 - p_{ijk})$.

- $R_i$ is the working correlation matrix for cluster $i$.

- The covariance of $\hat{\theta}$ can be obtained by the model-based estimator

$$(\sum_i D_i(\theta)^T V_i^{-1}(\alpha) D_i(\theta))^{-1}.$$ 

Hence, we can calculate the variance of $\hat{\beta}$. 
Design example

PPIUD study

- Outcome: unintended pregnancy within 1.5 years of the index birth
- 6 hospitals and 4 time periods
- $K=900$, cluster size $N = J \cdot K=3600$
- Baseline unintended pregnancy rate: 18.1%
- Expected unintended pregnancy rate after intervention
  - 14.5% – relative risk 0.8
  - 16.3% – relative risk 0.9
- ICC $\rho=0.022$
- Hypothetical time effects
  - No time effects $\delta = 0$
  - Negligible time effects $\delta = -0.01%$
  - 5% reduction of the baseline unintended pregnancy rate $\delta = -0.91%$
  - 10% reduction of the baseline unintended pregnancy rate $\delta = -1.81%$
### Design example

**Expected pregnancy rate**

<table>
<thead>
<tr>
<th>Time effects</th>
<th>( \delta = 0 )</th>
<th>( \delta = -0.01% )</th>
<th>( \delta = -0.91% )</th>
<th>( \delta = -1.81% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RR=0.8 )</td>
<td>( 1 )</td>
<td>( 0.908 )</td>
<td>( 0.976 )</td>
<td>( 0.981 )</td>
</tr>
<tr>
<td>( RR=0.9 )</td>
<td>( 0.976 )</td>
<td>( 0.480 )</td>
<td>( 0.494 )</td>
<td>( 0.506 )</td>
</tr>
</tbody>
</table>

**MLE (conditional)**

| \( RR=0.8 \) | \( 0.981 \)     | \( 0.494 \)     | \( 0.984 \)     | \( 0.506 \)     |
| \( RR=0.9 \) | \( 0.957 \)     | \( 0.449 \)     | \( 0.967 \)     | \( 0.474 \)     |

**GEE (marginal)**

| \( RR=0.8 \) | \( 0.935 \)     | \( 0.412 \)     | \( 0.935 \)     | \( 0.412 \)     |
| \( RR=0.9 \) | \( 0.935 \)     | \( 0.412 \)     | \( 0.935 \)     | \( 0.412 \)     |

**Hussey & Hughes**

| \( RR=0.8 \) | \( 0.935 \)     | \( 0.412 \)     | \( 0.935 \)     | \( 0.412 \)     |
| \( RR=0.9 \) | \( 0.935 \)     | \( 0.412 \)     | \( 0.935 \)     | \( 0.412 \)     |
Hussey and Hughes method is often inaccurate for binary outcomes. Power may be over- or under-estimated.

To improve this approximation, Zhou et al. proposed a conditional method, and Li et al. proposed a marginal method for binary outcomes, respectively.

The motivation for developing **swdpwr**:

These new methods had not yet been implemented in publicly available software such as SAS and R. Additionally, existing software for SWDs focused on limited settings and did not have accurate power calculations for binary outcomes.

We have developed user-friendly software **swdpwr** (SAS macro and R package) to implement power calculations for SWDs with both continuous and binary outcomes and incorporate as various settings as possible.

---

Two Statistical Models

- Two statistical models can be used to account for the intra-class correlations in SWDs: conditional models and marginal models.

- Conditional models are based on mixed effects models, which accommodate the intra-class correlations via latent random effects.

- Marginal models describe the population-averaged response values across cluster-periods, and are usually fitted with the generalized estimating equations (GEE).

- The interpretations of regression parameters can be different under these two models, with the important exception of the identity and log links when random effects and the covariates are independent, as is typically assumed.
Conditional Model - Binary Outcomes

For binary outcomes, consider a **cross-sectional** SWD which is modelled by a generalized linear mixed model (GLMM):

\[ g(p_{ijk}) = \mu + X_{ijk}\beta + \gamma_j + b_i. \]

- \( g(\cdot) \) can choose from: identity, log and logit links.
- \( b_i \sim N(0, \tau^2) \) is the random effect.
- The correlation structure for this case is \( \alpha_0 = \alpha_1 = \alpha_2. \)
- \( p_{ijk} = P(Y_{ijk} = 1|X_{ijk}, b_i) \) is the conditional probability of the outcome for each individual.
- Then we derive the likelihood of the observed outcomes based on the conditional probability under the specific link function. The calculation of the large-sample variance and power is then based on the theory of MLE.
- We also consider the case without time effects (all \( \gamma_t = 0 \)). The model is

\[ g(p_{ijk}) = \mu + X_{ijk}\beta + b_i. \]

The derivations of the likelihood formula and power are similar to the procedures with time effects.
Cluster randomized trials

- Intervention requires application at a cluster level
- Avoid contamination (e.g. sharing medication or health messages)
Software

Methods implemented:

- **Conditional model**\(^1\)
  - Only available for the cross-sectional design with \(\alpha_0 = \alpha_1\)
  - Three link functions: identity, log, logit
  - With time effects and without time effects
  - Intensive computation

- **Marginal model**\(^2\)
  - available for the cross-sectional design and cohort design with flexible covariance structures
  - Three link functions: identity, log, logit
  - With time effects and without time effects
  - Has closed-forms for power calculation.

---


The arguments of swdpower in the R package **swdpwr**

```r
swdpower(K, design, family = "binomial", model = "conditional",
         link = "identity", type = "cross-sectional",
         meanresponse_start = NA,
         meanresponse_end0 = meanresponse_start,
         meanresponse_end1 = NA,
         effectsize_beta = NA,
         sigma2 = 0,
         typeIerror = 0.05,
         alpha0 = 0.1,
         alpha1 = alpha0/2,
         alpha2 = NA)
```

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Number of individuals at each time period in a cluster, or the average cluster-period size under variable cluster-period sizes</td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>I*J dimensional data set that describes the study design, unequal allocation of sequences and only complete designs with no transition periods are allowed</td>
<td></td>
</tr>
<tr>
<td>family</td>
<td>Specify family=&quot;gaussian&quot; for continuous outcome, family=&quot;binomial&quot; for binary outcome</td>
<td>&quot;binomial&quot;</td>
</tr>
<tr>
<td>model</td>
<td>Specify model=&quot;conditional&quot; for conditional model. Model=&quot;marginal&quot; for marginal model</td>
<td>&quot;conditional&quot;</td>
</tr>
<tr>
<td>link</td>
<td>Choose link function from link=&quot;identity&quot;, Link=&quot;log&quot; and link=&quot;logit&quot;</td>
<td>&quot;identity&quot;</td>
</tr>
<tr>
<td>type</td>
<td>Specify type=&quot;cohort&quot; for closed cohort study. Type=&quot;cross-sectional&quot; for cross-sectional study</td>
<td>NA</td>
</tr>
<tr>
<td>meanresponse_start</td>
<td>The anticipated mean response in the control group at the start of the study</td>
<td>&quot;cross-sectional&quot;</td>
</tr>
<tr>
<td>meanresponse_end0</td>
<td>The anticipated mean response in the control group at the end of the study</td>
<td>meanresponse_start</td>
</tr>
<tr>
<td>meanresponse_end1</td>
<td>The anticipated mean response in the intervention group at the end of the study</td>
<td>NA</td>
</tr>
<tr>
<td>effectsize_beta</td>
<td>The anticipated effect size</td>
<td>NA</td>
</tr>
<tr>
<td>sigma2</td>
<td>Marginal variance of the outcome (only allowed if is continuous outcome)</td>
<td>0</td>
</tr>
<tr>
<td>typeIerror</td>
<td>Two-sided Type I error</td>
<td>0.05</td>
</tr>
<tr>
<td>alpha0</td>
<td>Within-period intracluster correlation $\alpha_0$</td>
<td>0.1</td>
</tr>
<tr>
<td>alpha1</td>
<td>Between-period intracluster correlation $\alpha_1$</td>
<td>Alpha0/2</td>
</tr>
<tr>
<td>alpha2</td>
<td>Within-individual correlation $\alpha_2$ (only allowed if type=&quot;cohort&quot;)</td>
<td></td>
</tr>
</tbody>
</table>
Example: PPIUD study

- **Outcome:** pregnancy within 1.5 years of the index birth
- 6 hospitals and 4 time periods
- \( K = 900 \), cluster size \( N = J \cdot K = 3600 \)
- Baseline pregnancy rate: 18.1%, and expected to be 16.3% after intervention
- ICC \( \rho = 0.022 \)
- Hypothetical time effects
  - No time effects \( \delta = 0 \)
  - Negligible time effects \( \delta = -0.01\% \)
Example: PPIUD study with no time effects

R> library("swdpwr")
R> data = matrix(c(rep(c(0, 1, 1, 1), times = 3),
+ rep(c(0, 0, 0, 1), times = 3)), nrow = 6, ncol = 4, byrow = TRUE)
R> swdpower(K = 900, design = data, family = "binomial", model = "marginal",
+ link = "logit", type = "cross-sectional", meanresponse_start = 0.181,
+ meanresponse_end0 = 0.181, meanresponse_end1 = 0.163,
+ typeIerror = 0.05, alpha0 = 0.022, alpha1 = 0.022)

This cross-sectional study has total sample size of 21600
Power for this scenario is 0.859 for the alternative hypothesis treatment effect
beta = -0.126 ( two-sided Type I error = 0.05 )
Example: PPIUD study with negligible time effects

R> library("swdpwr")
R> data = matrix(c(rep(c(0, 1, 1, 1), times = 3),
+ rep(c(0, 0, 0, 1), times = 3)), nrow = 6, ncol = 4, byrow = TRUE)
R> swdpower(K = 900, design = data, family = "binomial", model = "marginal",
+ link = "logit", type = "cross-sectional", meanresponse_start = 0.181,
+ meanresponse_end0 = 0.1809, meanresponse_end1 = 0.163,
+ typeIerror = 0.05, alpha0 = 0.022, alpha1 = 0.022)

This cross-sectional study has total sample size of 21600
Power for this scenario is 0.423 for the alternative hypothesis treatment effect 
 beta = -0.126 (two-sided Type I error = 0.05)
Conclusion

- The Hussey and Hughes method is inaccurate for power calculation with binary outcomes. Power may be over- or under-estimated.

- Two methods using mixed effects models and GEEs can be used to calculate power of SWDs with binary outcomes.

- We have developed user-friendly software `swdpwr`, including a SAS macro and an R package, to implement power calculation methods for SWDs.
References


Contact me at xin.zhou@yale.edu for questions, comments, and suggestions.