

## Adjusted proportion difference and confidence interval in stratified randomized trials

Yeonhee Kim, Gilead Sciences, Seattle, WA

Seunghyun Won, University of Pittsburgh, Pittsburgh, PA

### ABSTRACT

Stratified randomization is widely used in clinical trials to achieve treatment balance across strata. In the analysis, investigators are often interested in the estimation of common treatment effect adjusting for stratification factors. We developed the macro for stratified analysis when the primary endpoint is a difference in proportions between two treatment groups. The %strataCI macro estimates proportion difference using three different weighting schemes and calculates Wald and Newcombe confidence intervals.

Keywords: Stratified analysis, weight-adjusted proportion, Minimum risk weight, Wald, Newcombe confidence interval

### INTRODUCTION

Stratified randomization is often used when treatment outcome is presumed to be affected by baseline characteristics such as gender and disease severity. Another example of stratification can be found in multicenter trials in which patients are randomly assigned to the treatments within each center. In the analysis, treatment effects are not only assessed separately by each center but also combined after adjusting for center effect. This paper focuses on the stratified analysis for binary outcomes (success or failure) when the primary endpoint is a proportion difference between two treatment groups. Starting from the stratum-specific proportion difference, we consider the point estimate and corresponding confidence interval for the strata-adjusted proportion difference. The adjusted effect is obtained by weighted average of stratum specific rate differences. For treatment comparison, we present Wald type confidence interval as well as stratified Newcombe confidence interval.

### METHOD

#### STRATUM-SPECIFIC PROPORTION DIFFERENCE

Let  $p_{ij}$  is the proportion of success in treatment  $i$  ( $i = 1, 2$ ) and stratum  $j$  ( $j = 1, \dots, s$ ), then the stratum-specific proportion difference is defined as  $d_j = p_{1j} - p_{2j}$ . In the below 2x2 contingency table for the  $j^{\text{th}}$  stratum, one can estimate the proportion of success  $\hat{p}_{ij} = x_{ij}/n_{ij}$ , and the stratum-specific proportion difference  $\hat{d}_j = \hat{p}_{1j} - \hat{p}_{2j} = x_{1j}/n_{1j} - x_{2j}/n_{2j}$ .

Stratum j	Treatment 1	Treatment 2
Failure	$x_{1j}$	$x_{2j}$
No failure	$n_{1j} - x_{1j}$	$n_{2j} - x_{2j}$
Total	$n_{1j}$	$n_{2j}$

Table 1 2x2 contingency table ( $j^{\text{th}}$  stratum)

#### ADJUSTED PROPORTION DIFFERENCE

Strata-adjusted proportion difference ( $d_{adj}$ ) can be obtained by weighted average of stratum-specific proportion differences, that is,  $d_{adj} = \sum_{j=1}^s w_j d_j$ , where  $w_j$  is the weight assigned to stratum  $j$ . Several weighting strategies have been developed such as Cochran-Mantel-Haenszel, inverse of variance and minimum risk weights (Mehrotra and Railkar, 2000).

Cochran-Mantel-Haenszel weight is given by

$$w_j = \frac{\left( \frac{n_{1j}n_{2j}}{n_{1j} + n_{2j}} \right)}{\sum_{j=1}^s \left( \frac{n_{1j}n_{2j}}{n_{1j} + n_{2j}} \right)}.$$

It assigns more weight to the stratum with large sample size than the stratum of small size. The Cochran-Mantel-Haenszel weight is especially powerful to detect treatment difference when the odds ratio  $\frac{\hat{p}_{1j}(1-\hat{p}_{2j})}{\hat{p}_{2j}(1-\hat{p}_{1j})}$  is constant across strata. While the Cochran-Mantel-Haenszel weight does not take variability of  $\hat{d}_j$  into account, inverse of variance weight considers heterogeneity between strata. The inverse of variance weight has the form

$$w_j = \frac{\left( \frac{p_{1j}(1-p_{1j})}{n_{1j}} + \frac{p_{2j}(1-p_{2j})}{n_{2j}} \right)^{-1}}{\sum_{j=1}^s \left( \frac{p_{1j}(1-p_{1j})}{n_{1j}} + \frac{p_{2j}(1-p_{2j})}{n_{2j}} \right)^{-1}},$$

which implies that the stratum with high precision for the risk difference (i.e. smaller variance of  $\hat{d}_j$ ) has more weights.

Minimum risk weight is driven from the risk function so that the squared error loss is minimized. The minimum risk weight is given as

$$w_j = \frac{\beta_j}{\sum_{k=1}^s V(\hat{d}_k)^{-1}} - \left[ \frac{\alpha_j V(\hat{d}_j)^{-1}}{\sum_{k=1}^s V(\hat{d}_k)^{-1} + \sum_{k=1}^s \alpha_k d_k V(\hat{d}_k)^{-1}} \right] * \left[ \frac{\sum_{k=1}^s d_k \beta_k}{\sum_{k=1}^s V(\hat{d}_k)^{-1}} \right],$$

where  $\alpha_j = d_j \sum_{k=1}^s V(\hat{d}_k)^{-1} - \sum_{k=1}^s d_k \hat{V}(\hat{d}_k)^{-1}$ ,  $\beta_j = V(\hat{d}_j)^{-1} \left( 1 + \alpha_j \sum_{k=1}^s \frac{d_k(n_{1k}+n_{2k})}{\sum_{m=1}^s (n_{1m}+n_{2m})} \right)$ ,

and  $\hat{V}(\hat{d}_j) = \frac{\hat{p}_{1j}(1-\hat{p}_{1j})}{n_{1j}} + \frac{\hat{p}_{2j}(1-\hat{p}_{2j})}{n_{2j}}$ .

## CONFIDENCE INTERVAL FOR ADJUSTED PROPORTION DIFFERENCE

Confidence interval is frequently reported along with the point estimate in comparative studies. If the confidence interval for  $\hat{d}_{adj}$  does not include value of zero, difference between two groups are considered statistically different. The Wald-type 100 x (1- $\alpha$ )% confidence interval is constructed as

$$\hat{d}_{adj} \pm z_{\alpha/2} \sqrt{\sum_{j=1}^s \hat{w}_j^2 \left( \frac{\hat{p}_{1j}(1-\hat{p}_{1j})}{n_{1j}} + \frac{\hat{p}_{2j}(1-\hat{p}_{2j})}{n_{2j}} \right)},$$

where  $z_{\alpha/2}$  the 100 x (1 -  $\alpha/2$ )<sup>th</sup> percentile of the standard normal distribution. Although Wald-type confidence interval has been widely used in practice due to computational simplicity, coverage probability is not always good, especially when  $\hat{d}_{adj}$  approaches to 0 or 1 (Agresti and Coull, 1998; Agresti and Caffo, 2000). Stratified Newcombe confidence interval is shown to give better coverage probabilities. The stratified Newcombe confidence interval proposed by Yan and Su (2010) has the form

$$\left[ \hat{d}_{adj} - z_{\frac{\alpha}{2}} \sqrt{\sum_{j=1}^s \frac{\hat{w}_j^2}{n_{1j}} L_1(1-L_1) + \sum_{j=1}^s \frac{\hat{w}_j^2}{n_{2j}} U_2(1-U_2)}, \hat{d}_{adj} + z_{\frac{\alpha}{2}} \sqrt{\sum_{j=1}^s \frac{\hat{w}_j^2}{n_{2j}} L_2(1-L_2) + \sum_{j=1}^s \frac{\hat{w}_j^2}{n_{1j}} U_1(1-U_1)} \right],$$

where each  $(L_i, U_i)$  is the Wilson confidence interval for  $\hat{p}_i = \sum_{j=1}^s w_j \hat{p}_{ij}$ ,  $i = 1, 2; j = 1, 2 \dots s$ . Each  $(L_i, U_i)$  is computed as

$$(L_i, U_i) = \sum_{j=1}^s \hat{w}_j \left( \frac{\hat{p}_{ij} + \frac{z_{\gamma i}^2}{2n_j}}{1 + \frac{z_{\gamma i}^2}{n_j}} \pm \frac{z_{\gamma i}}{1 + \frac{z_{\gamma i}^2}{n_j}} \sqrt{\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n_j} + \frac{z_{\gamma i}^2}{4n_j^2}} \right), \text{ where } z_{\gamma i} = z_{\alpha/2} \left( \frac{\sqrt{\sum_{j=1}^s \hat{w}_j^2 \left( \frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n_j} \right)}}{\sum_{j=1}^s \hat{w}_j \sqrt{\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n_j}}} \right).$$

## SAS MACRO IMPLEMENTATION

For demonstration purpose, we use the following example dataset that contains subject id, response (success = 1, failure = 0), treatment group (1, 2), and stratum number (1, 2, 3).

id	treatment	strata	response
1	1	2	1
2	2	2	0
3	1	2	0
4	1	2	0
5	2	3	0
6	1	1	1
7	1	2	1
8	1	3	0
9	1	1	0
10	1	1	1
11	1	2	0
12	1	3	0

**Display 1. Example dataset (first 12 observations)**

To execute the macro %strataCI, users need to define the SAS dataset that will be analyzed and the variables representing treatment group, strata, response and type 1 error.

```
%strataCI(data= , treatment= , strata= , response= , level= );
```

Variable	Description
data	Input SAS dataset name <i>The dataset should contain variables for subject id, treatment, strata and response.</i>
treatment	Variable name for treatment in the SAS dataset <i>Data type for treatment is numeric, e.g., 1 for Treatment and 0 for control</i>
strata	Variable name for strata in the SAS dataset <i>Data type for strata is numeric, e.g., 1, 2, 3,...</i>
response	Variable name for response in the SAS dataset <i>Data type for response is numeric, e.g., 1 for success and 0 for failure</i>
level	Type I error

**Table 2 Description of input variables**

The %strataCI macro generates summary of strata-adjusted proportion difference along with stratum-specific proportion difference using PROC FREQ and PROC IML. For comparison, the program also calculates confidence interval for crude proportion difference. The stratum-specific summary output (Table 3) includes sample sizes ( $n_{1j}$  and  $n_{2j}$ ), proportions of success ( $p_{1j}$  and  $p_{2j}$ ), proportion difference between two treatment groups ( $d_j$ ) and weights ( $w_j$ ) assigned in each stratum  $j$  ( $j = 1, 2, 3$ ). In the summary output, for example, the highest weight is assigned to the second stratum.

Stratum-specific summary									
	Strata	Sample size in Trt1	Sample size in Trt2	Proportion in Trt1	Proportion in Trt2	Proportion difference (Trt1 - Trt2)	CMH weight	Inverse of Variance weight	Minimum risk weight
<b>ROW1</b>	1	19	24	0.8421053	0.2083333	0.6337719	0.2132632	0.3028324	0.2308761
<b>ROW2</b>	2	49	55	0.4285714	0.4	0.0285714	0.5211286	0.4486808	0.5270501
<b>ROW3</b>	3	28	25	0.6428571	0.32	0.3228571	0.2656083	0.2484868	0.2420738

**Table 3 Output of stratum-specific summary**

The strata-adjusted summary output (Table 4) presents adjusted proportion difference ( $d_{adj}$ ), stratified Wald confidence interval and stratified Newcombe confidence interval by each weighting scheme.

Adjusted summary					
	Adjusted proportion difference	Wald CI (lower)	Wald CI (upper)	Newcombe CI (lower)	Newcombe CI (upper)
<b>CMH</b>	0.236	0.106	0.365	0.098	0.363
<b>Inverse of Variance</b>	0.285	0.158	0.412	0.144	0.412
<b>Minimum risk</b>	0.240	0.111	0.369	0.101	0.366

**Table 4 Output of strata-adjusted summary**

The confidence interval derived from crude proportion difference is shown in Table 5.

Unadjusted summary	
Crude CI (lower)	Crude CI (upper)
0.102	0.371

**Table 5 Output of crude summary**

## CONCLUSION

The paper describes SAS macro to compute strata-adjusted proportion difference and its confidence interval. Users can compare results and choose a proper weighting scheme out of Cochran-Mantel-Haenszel, inverse of variance and minimum risk weights. Difference in treatment effects is assessed using stratified Wald or Newcombe confidence interval. The macro can be easily applied for multi-center study, stratified randomization trials, studies with confounding factors, and meta analysis.

## REFERENCES

- Agresti A. and Caffo B. (2000). "Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures." *The American statistician* 54 (4), 280-288.
- Agresti A. and Coull B. (1998). "Approximation is better than "Exact" for interval estimation of binomial proportions." *The American statistician* 52 (2), 119-126.
- Mehrotra D. and Railkar R. (2000). "Minimum risk weights for comparing treatments in stratified binomial trials." *Statistics in Medicine* 19, 811-825.
- Yan X. and Su X.G. (2010). "Stratified Wilson and Newcombe confidence intervals for multiple binomial proportions." *Statistics in Biopharmaceutical Research* 2(3), 329-335.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Yeonhee Kim  
 Gilead Sciences  
 199 East Blaine Street  
 Seattle, WA 98102  
 Yeonhee.Kim@gilead.com

Seunghyun Won  
 Department of Biostatistics  
 University of Pittsburgh  
 130 DeSoto Street  
 Pittsburgh, PA 15261  
 sew53@pitt.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

## APPENDIX

```
%macro stratCI(data=, treatment=, strata=, response=, level=);

  %if %sysfunc(exist(&data))=0 %then %do;
    %put ' ERROR: Input dataset not specified';
    %goto exit;
  %end;
  %if %length(&treatment) = 0 %then %do;
    %put ' ERROR: Treatment group must be specified';
    %goto exit;
  %end;
  %if %length(&strata) = 0 %then %do;
    %put ' ERROR: Strata must be specified';
    %goto exit;
  %end;
  %if %length(&response) = 0 %then %do;
    %put ' ERROR: Response must be specified';
    %goto exit;
  %end;

  proc sql noprint;
  create table temp as
    select
      &treatment as treatment,
      &strata as strata,
      &response as response
    from &data;
  quit;

  %let alpha= 1-&level/2;

  proc sort data=temp;
    by strata treatment;
  run;

  proc freq data=temp noprint;
    by strata;
    tables treatment/riskdiff nopercnt nocol norow out=freq_strata;
  run;

  proc freq data=temp;
    by strata;
    tables treatment*response/riskdiff nopercnt nocol out=freq;
    ods output RiskDiffCol2 = riskstrata;
  run;

  * Total Frequency;
  proc transpose data=freq_strata out=freq_stratal (drop=_NAME_ _LABEL_) prefix=n_trt;
    by strata;
    var count;
  run;

  * Proportion;
  proc transpose data=riskstrata out=riskstratal (drop=_NAME_) prefix=prop;
    where Row in ('Row 1','Row 2');
    by strata;
    var Risk;
  run;

  * Risk difference;
  data riskstrata2;
    set riskstrata(rename=(risk=r_diff));
    if Row = 'Difference';
    keep r_diff ASE strata;
  run;

  * Merge;
  data ready;
    merge freq_stratal riskstratal riskstrata2;
    by strata;
  run;

  *STRATA-ADJUSTED RISK DIFFERENCE;
  data ready1;
    set ready;
    *calculate weight;

    CMH_part = (n_trt1*n_trt2)/(n_trt1+n_trt2);
    invar part =1/ASE**2;
```

```
min_part1= (n_trt1+n_trt2);
min_part2= min_part1*r_diff;
min_part3= invar_part*r_diff;
run;

proc sql;
  create table ready2 as
  select strata, sum(CMH_part) as CMH_sum,
  sum(invar_part) as invar_sum,
  sum(min_part1) as min_sum1,
  sum(min_part2) as min_sum2,
  sum(min_part3) as min_sum3
  from ready1;
quit;

data ready3;
  merge ready1 ready2;
  by strata;
  min_alpha = r_diff*invar_sum - min_sum3;
  min_beta = invar_part * (1 + min_alpha*min_sum2/min_sum1);
  min_part4=r_diff*min_beta;
  min_part5 = r_diff*min_alpha*invar_part;
run;

proc sql;
  create table ready3_1 as
  select *,
  sum(min_part4) as min_sum4,
  sum(min_part5) as min_sum5
  from ready3;
quit;

data weights;
  set ready3_1;
  CMH_weight = CMH_part / CMH_sum;
  INVAR_weight = invar_part / invar_sum;
  MIN_weight = min_beta/invar_sum
  - (min_alpha*invar_part/(invar_sum + min_sum5))*(min_sum4/ invar_sum);
run;

/* all weights are calculated. */

proc iml;
  use weights;
  read all;
  read all var {CMH_weight invar_weight min_weight} into w;

  n=n_trt1 || n_trt2;
  n_strata=nrow(n);
  totaln=sum(n);
  p= prop1 || prop2;

  /* unadjusted CI :Pool */

  p1_1 = n_trt1 # prop1; p2_1 = n_trt2 # prop2;
  n1=n_trt1[+,]; n2=n_trt2[+,];
  p1=p1_1[+,]/n1; p2=p2_1[+,]/n2;
  Crude_Lower = (p1-p2) - probit(&alpha)# sqrt(p1 # (1-p1)/n1 + p2 # (1-p2)/n2);
  Crude_Upper = (p1-p2) + probit(&alpha)# sqrt(p1 # (1-p1)/n1 + p2 # (1-p2)/n2);

  /* start to calculate Wald CI */

  adj_rdiff= w # r_diff;
  wald_part1= adj_rdiff[+,];
  wald_var = 1/invar_part;      sq_w = w ## 2;

  adj_var = sq_w # wald_var;
  wald_part2 = adj_var[+,];

  Wald_L = wald_part1 - probit(&alpha)*sqrt(wald_part2);
  Wald_U = wald_part1 + probit(&alpha) *sqrt(wald_part2);

  Wald_Lower=Wald_L`;
  Wald_Upper=Wald_U`;

  /* end of Wald CI */
```

```

/* start to calculate Newcombe */
* 1. calculate z_gamma;

partz_a1 = J(n_strata,3,1);partz_a2 =J(n_strata,3,1);
partz_b1 = J(n_strata,3,1);partz_b2 = J(n_strata,3,1);

do i=1 to n_strata;
  do j=1 to 3;
    partz_a1[i,j] = (sq_w[i,j]/n_trt1[i,1]) # (p[i,1] # (1-p[i,1]));
    partz_a2[i,j] = (sq_w[i,j]/n_trt2[i,1]) # (p[i,2] # (1-p[i,2]));
    partz_b1[i,j] = w[i,j] # sqrt(p[i,1] # (1-p[i,1])/n_trt1[i,1]);
    partz_b2[i,j] = w[i,j] # sqrt(p[i,2] # (1-p[i,2])/n_trt2[i,1]);
  end;
end;

z_value1=(probit(&alpha) * sqrt(partz_a1[+,]) / partz_b1[+,])`;
z_value2=(probit(&alpha) * sqrt(partz_a2[+,]) / partz_b2[+,])`;

* 2. calculate L1 L2 U1 U2 based on stratified Wilson CI;

wilson_part1a= J(n_strata,3,1); wilson_part1b= J(n_strata,3,1);
wilson_part2a= J(n_strata,3,1); wilson_part2b= J(n_strata,3,1);

do i=1 to n_strata;
  do j=1 to 3;
wilson_part1a[i,j]=(n[i,1] # p[i,1] + 0.5 # (Z_value1[j,1] ## 2)) / (n[i,1] + Z_value1[j,1] ## 2);
wilson_part1b[i,j]=(n[i,1] # Z_value1[j,1])/(n[i,1]+Z_value1[j,1]##2)
# (sqrt(4 # n[i,1] # p[i,1] # (1-p[i,1]) + Z_value1[j,1] ##2)/(2 # n[i,1]) );
wilson_part2a[i,j]=(n[i,2] # p[i,2] + 0.5 # (Z_value2[j,1] ## 2)) / (n[i,2] + Z_value2[j,1] ## 2);
wilson_part2b[i,j]=(n[i,2] # Z_value2[j,1])/(n[i,2]+Z_value2[j,1]##2)
# (sqrt(4 # n[i,2] # p[i,2] # (1-p[i,2]) + Z_value2[j,1] ##2)/(2 # n[i,2]) );
  end;
end;

wilson_L1 = w # ( wilson_part1a - wilson_part1b);
wilson_U1 = w # ( wilson_part1a + wilson_part1b);

wilson_L2 = w # ( wilson_part2a - wilson_part2b);
wilson_U2 = w # ( wilson_part2a + wilson_part2b);

newcombe_L1 = wilson_L1[+,];
newcombe_U1 = wilson_U1[+,];

newcombe_L2 =wilson_L2[+,];
newcombe_U2 =wilson_U2[+,];

* 3. Calculate lambda1 and 2;

lam1= sq_w / n_trt1; lam2 = sq_w / n_trt2;
lambda1 = lam1[+,]; lambda2 = lam2[+,];

* 4. calculate lower/upper limit for newcombe CI;

NEWCOMBE_L = wald_part1 - probit(&alpha) # sqrt(lambda1 # newcombe_L1
# (1- newcombe_L1)+lambda2 # newcombe_U2 # (1-newcombe_U2));
NEWCOMBE_U = wald_part1 + probit(&alpha) # sqrt(lambda2 # newcombe_L2
# (1-newcombe_L2)+lambda1 # newcombe_U1 # (1-newcombe_U1));

Newcombe_Lower=newcombe_L`; Newcombe_Upper=newcombe_U`;

/* end of Newcombe CI */

Stratum_nm = {'Strata','Sample size in Trt1','Sample size in Trt2','Proportion in Trt1',
'Proportion in Trt2','Proportion difference (Trt1 - Trt2)','CMH weight',
'Inverse of Variance weight', 'Minimum risk weight' };
Adj_nm = {"Adjusted proportion difference", "Wald CI (lower)", "Wald CI (upper)",
"Newcombe CI (lower)", "Newcombe CI (upper)"};
Weights= { "CMH", "Inverse of Variance", "Minimum risk"};
CI = {"Crude CI (lower)", "Crude CI (upper)"};

Strata_sum = (1:n_strata)`||n_trt1|| n_trt2 || prop1 || prop2 || r_diff || CMH_weight ||
invar_weight || min_weight;
Adj_sum = wald_part1`||wald_Lower || Wald_Upper ||newcombe_Lower || newcombe_Upper ;
Crude_CI = crude_Lower` || crude_Upper`;

file print;
put @35 "Strata-adjusted confidence interval"//;
put @35 "Summary for dataset "/;
put @30 "Sample size " @75 totaln /
@30 "Number of strata " @75 n_strata;

```

< Adjusted proportion difference and confidence interval in stratified randomized trials >, continued

```
print Strata_sum [colname=stratum_nm label= "Stratum-specific summary" ];
print crude_CI[colname=CI label= "Unadjusted summary" format=6.3];
print Adj_sum [rowname=weights colname=Adj_nm label= "Adjusted summary" format=6.3];

quit;

%exit:
%mend stratCI;

%stratCI(data=example, treatment=treatment, strata=strata, response=response, level=0.05);
```