

Mantel-Haenszel法により調整したリスク差の信頼区間に関する一考察

○佐々木 英麿、舟尾 暢男

武田薬品工業株式会社 日本開発センター 生物統計室

A Consideration of the Confidence Intervals for Mantel-Haenszel Estimate of Common Risk Difference

Emma Sasaki, Nobuo Funao

Takeda Pharmaceutical Company, Ltd.

要旨

Mantel-Haenszel法により調整したリスク差に関する信頼区間の算出方法(Greenland信頼区間、Sato信頼区間、及びNewcombe信頼区間)を紹介し、各信頼区間の被覆確率をシミュレーションにより確認することで性能評価を行う。

キーワード : Mantel-Haenszel 法、調整済みリスク差、信頼区間、被覆確率、FREQ プロシジャ、STDRATE プロシジャ

1. はじめに

2 値応答変数を主要評価項目とした 2 群の並行群間比較試験では、主解析として割合の群間差 (以下、リスク差) の評価を行い、感度分析として背景因子で調整した解析を実施するのが一般的である。また、治験薬の割付において層別割付を行った場合、主解析として層別因子で調整した解析を計画することがあるが、その際、Cochran-Mantel-Haenszel 検定 (以下、CMH 検定) 等を用いて群間比較を実施するとともに、Mantel-Haenszel 法により調整したリスク差 (以下、調整済みリスク差) に関する推定量を算出することが一般的である。いずれの場合においても、調整済みリスク差を算出する場合は、対応する信頼区間を合わせて算出することとなるが、信頼区間の算出方法についてはいくつか選択肢がある。

SAS では、リスク差に関する様々な信頼区間の算出が可能となっている。SAS V9.3 で算出可能な信頼区間については飯塚、浜田 (2013) が性能評価を行っており、SAS V9.4 で新たに追加された信頼区間については飯塚、魚住、浜田 (2014) が性能評価を行い、いずれも Newcombe スコアに基づく信頼区間を推奨しているが、調整済みリスク差に関する信頼区間の考察は、本ユーザー総会ではなされていない。

本稿では、FREQ プロシジャに加え、SAS V9.3 で新たに追加された STDRATE プロシジャを用い、調整済みリスク差に対する信頼区間の算出方法 (Greenland 信頼区間、Sato 信頼区間、及び Newcombe 信頼区間) を紹介し、これらのプロシジャを用いて調整済みリスク差の信頼区間を算出する際の留意点を示すとともに、それぞれの信頼区間の被覆確率をシミュレーションにより確認し、性能評価を行う。

2. Mantel-Haenszel 法による調整済みリスク差の算出

まずは、表 2.1 の分割表を考える。

表 2.1 2 × 2 分割表

投与群	イベントの有無		合計
	あり	なし	
群 1	x_1	$n_1 - x_1$	n_1
群 2	x_2	$n_2 - x_2$	n_2
			n_+

このとき、群 j の被験者のイベント発生割合（リスク）を p_j とすると（ $j = 1, 2$ ）、各群のイベント発生数に関する確率変数 X_j は二項分布 $B(n_j, p_j)$ に従うが、漸近的には正規分布 $N(n_j p_j, n_j p_j (1 - p_j))$ に従う。よって、各群の割合（リスク）に関する確率変数 Y_j は以下のように示すことができる。

$$Y_j = \frac{X_j}{n_j} \sim N\left(p_j, \frac{p_j(1-p_j)}{n_j}\right) \quad (1)$$

2 群の並行群間比較試験において、群 1 と群 2 の割合（リスク）に関する確率変数 Y_1 と Y_2 はそれぞれ独立した正規分布に従うため、群 1 と群 2 のリスク差（ $Y_1 - Y_2$ ）の分布は以下の通りとなる。

$$D = Y_1 - Y_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right) \quad (2)$$

次に調整済みリスク差を考えるため、表 2.2 の分割表を考える。

表 2.2 層 i の 2 × 2 分割表（ $i = 1, \dots, k$ ）

投与群	イベントの有無		合計
	あり	なし	
群 1	x_{i1}	$n_{i1} - x_{i1}$	n_{i1}
群 2	x_{i2}	$n_{i2} - x_{i2}$	n_{i2}
			n_{i+}

k 個の層を有する、ある層別項目を考えた場合、層 i の群 j における被験者のイベント発生割合（リスク）を p_{ij} とすると（ $i = 1, \dots, k; j = 1, 2$ ）、層 i における各群のイベント発生数に関する確率変数 X_{ij} は二項分布 $B(n_{ij}, p_{ij})$ に従い、漸近的には正規分布 $N(n_{ij} p_{ij}, n_{ij} p_{ij} (1 - p_{ij}))$ に従う。よって、層 i における各群の割合（リスク）に関する確率変数 Y_{ij} は以下の分布に従う。

$$Y_{ij} = \frac{X_{ij}}{n_{ij}} \sim N\left(p_{ij}, \frac{p_{ij}(1-p_{ij})}{n_{ij}}\right) \quad (3)$$

よって、層 i のリスク差（ $Y_{i1} - Y_{i2}$ ）の分布は以下の通りとなる。

$$D_i = Y_{i1} - Y_{i2} \sim N\left(p_{i1} - p_{i2}, \frac{p_{i1}(1-p_{i1})}{n_{i1}} + \frac{p_{i2}(1-p_{i2})}{n_{i2}}\right) \quad (4)$$

式 (4) より、Mantel-Haenszel 型の重み $w_i = n_{i1}n_{i2}/(n_{i1} + n_{i2})$ を用いて重み付き推定量を求めると、Mantel-Haenszel 法により調整したリスク差（調整済みリスク差）が得られる。

$$\widehat{RD}_{MH} = \frac{\sum_{i=1}^k w_i * D_i}{\sum_{i=1}^k w_i} = \frac{\sum_{i=1}^k (x_{i1}n_{i2} - x_{i2}n_{i1})/n_{i+}}{\sum_{i=1}^k n_{i1}n_{i2}/n_{i+}} \quad (5)$$

次項では、調整済みリスク差の信頼区間を考える。

3. 調整済みリスク差の信頼区間の算出方法

3.1 Greenland 信頼区間

調整済みリスク差の信頼区間を算出するにあたって、まず、Greenland and Robins (1985) の方法に基づいた信頼区間（以下、Greenland 信頼区間）を考える。本方法では、式 (6) に従って調整済みリスク差の分散を計算し、この分散に基づいて信頼区間を算出する。

$$\widehat{\text{var}}\left(\frac{\sum_{i=1}^k w_i * D_i}{\sum_{i=1}^k w_i}\right) = \frac{\sum_{i=1}^k w_i^2 \left(\frac{x_{i1}(n_{i1} - x_{i1})}{n_{i1}^3} + \frac{x_{i2}(n_{i2} - x_{i2})}{n_{i2}^3} \right)}{(\sum_{i=1}^k w_i)^2} \quad (6)$$

Sato (1989) では、Greenland 信頼区間はデータが疎な場合に一致性を持たないことを指摘している。これに対して Greenland & Robins は指摘を認めており、Greenland 信頼区間は層の例数が多い場合にのみ一致性を持つと訂正をしている。一方、Rothman, Lash, and Greenland (2012b) では、各層の各群の分母が全て 2 以上であれば、Greenland 信頼区間を用いることが可能だとしており、Greenland 信頼区間を使用する是非の判断が難しい状況である。今回の検討ではこの点も含めて確認を行う。

3.2 Sato 信頼区間

Sato (1989) では、Greenland and Robins (1985) の方法を改善し、層の数が多くデータが疎な場合でも妥当となるような調整済みリスク差の分散の推定方法を提案している。Sato (1989) の方法に基づいた信頼区間（以下、Sato 信頼区間）は式 (7) に従って調整済みリスク差の分散を計算し、この分散に基づいて信頼区間を算出する。この信頼区間は、すべての場合のデータにおいて妥当であることが Rothman, Lash, and Greenland (2012b) に記されている。

$$\widehat{\text{var}}^*(\widehat{RD}_{MH}) = (\widehat{RD}_{MH} \sum_{i=1}^k P_i + \sum_{i=1}^k Q_i) / (\sum_{i=1}^k w_i)^2 \quad (7)$$

このとき、 P_i 及び Q_i はそれぞれ

$$\begin{aligned} P_i &= \left(n_{i1}^2 x_{i2} - n_{i2}^2 x_{i1} + \frac{n_{i1}n_{i2}(n_{i2} - n_{i1})}{2} \right) / (n_{i1} + n_{i2})^2, \\ Q_i &= (x_{i1}(n_{i2} - x_{i2}) - x_{i2}(n_{i1} - x_{i1})) / 2(n_{i1} + n_{i2}) \end{aligned} \quad (8)$$

となる。

3.3 Newcombe 信頼区間

Yan and Su (2010) の方法で stratified Wilson 信頼区間に基づいて構築される stratified Newcombe 信頼区間（以下、Newcombe 信頼区間）では、調整済みリスク差 (\widehat{RD}_{MH}) の $100(1 - \alpha)\%$ 信頼下限と信頼上限 (L, U) は以下の通りとなる。

$$\begin{aligned}
 L &= \widehat{RD}_{MH} - z_{\alpha/2} \sqrt{\lambda_1 L_1 (1 - L_1) + \lambda_2 U_2 (1 - U_2)}, \\
 U &= \widehat{RD}_{MH} + z_{\alpha/2} \sqrt{\lambda_1 U_1 (1 - U_1) + \lambda_2 L_2 (1 - L_2)}
 \end{aligned}
 \tag{9}$$

このとき、 λ_1 と λ_2 は

$$\lambda_1 = \sum_{i=1}^k w_i^2 / n_{i1}, \quad \lambda_2 = \sum_{i=1}^k w_i^2 / n_{i2}
 \tag{10}$$

であり、 L_1, U_1 は群 1 の割合（リスク）の stratified Wilson score 信頼上限と信頼下限を表し、 L_2, U_2 は群 2 の割合（リスク）の stratified Wilson score 信頼上限と信頼下限を表す。冒頭でも述べたが、層を考慮しないリスク差の信頼区間の性能評価の結果、飯塚、浜田 (2013) 及び飯塚、魚住、浜田 (2014) は、いずれも Newcombe スコアに基づく信頼区間を推奨している。

3.4 それぞれの信頼区間を算出する SAS プログラム

それぞれの信頼区間を算出する例を紹介するため、表 3.1 を考える。なお、下記サンプルデータは Rothman (2012a) から引用している。

表 3.1 サンプルデータ

投与群	層 1: イベントの有無		合計	投与群	層 2: イベントの有無		合計
	あり	なし			あり	なし	
群 1	8	98	106	群 1	22	76	98
群 2	5	115	120	群 2	16	69	85
			226				183

まず、Greenland 信頼区間は STDRATE プロシジャにより算出が可能である。Greenland 信頼区間の算出プログラムをプログラム 3.1 に、実行結果を出力結果 3.1 に示す。

プログラム 3.1

```

data SAMPLEDATA1 ;
  input TREAT STRATA X TOTAL ;
  cards ;
    1 1 8 106
    1 2 22 98
    2 1 5 120
    2 2 16 85
  ;
run ;

ods graphics on;
proc stdrate data=SAMPLEDATA1
  method=mh
  stat =risk
  effect=diff
  plots =all ;
  population group=TREAT event=X total=TOTAL ;
  strata STRATA / order=internal stats(c1=normal) effect ;
run;
ods graphics off;

```

出力結果 3.1 (抜粋)

Risk Effect Estimates							
-----TREAT-----		Risk	95% Normal		Standard	Z	Pr > Z
1	2	Difference	Confidence Limits		Error		
0.1421	0.1072	0.0349	-0.027619	0.097419	0.0319	1.09	0.2739

次に、Sato 信頼区間及び Newcombe 信頼区間は FREQ プロシジャにより算出が可能である。Sato 信頼区間は riskdiff オプションへ「common column=1」を加えることによって算出が可能であり、Newcombe 信頼区間はさらに「cl=newcombe」を追記することによって算出可能となる。Sato 信頼区間及び Newcombe 信頼区間の算出プログラムをプログラム 3.2 に、実行結果を出力結果 3.2 に示す。なお、Greenland 信頼区間を STDRATE プロシジャで算出する際とは、データセットの形が異なることに留意されたい。

プログラム 3.2

```
data SAMPLEDATA2 ;
  input STRATA TREAT X N ;
  cards ;
  1 1 1 8
  1 1 2 98
  1 2 1 5
  1 2 2 115
  2 1 1 22
  2 1 2 76
  2 2 1 16
  2 2 2 69
  ;
run ;

proc freq data=SAMPLEDATA2 ;
  table STRATA*TREAT*X / riskdiff(common column=1 cl=newcombe) nocol nopercnt ;
  weight N ;
run ;
```

出力結果 3.2 (抜粋)

手法	共通比率 (リスク) 差		95% 信頼限界	
	値	標準誤差		
Mantel-Haenszel	0.0349	0.0319	-0.0276	0.0974
Newcombe	0.0349		-0.0302	0.1000
要約スコア	0.0372	0.0299	-0.0215	0.0959

列 1 (X = 1)

3.5 プログラムに関する注意点

前項では、調整済みリスク差とその信頼区間を STDRATE プロシジャ又は FREQ プロシジャを用いて算出する例を紹介したが、データの状態によっては、いくつかの信頼区間が算出されない場合があるため注意が必要である。

まず、表 3.2 のように、全ての層において、「全ての群が 0%」又は「全ての群が 100%」となった場合、STDRATE プロシジャ又は FREQ プロシジャを実施してもいずれの信頼区間も算出されない。すなわち、表 3.2 のような場合は全ての信頼区間が算出できない。

表 3.2 全ての信頼区間が算出できない場合

投与群	層 1: イベントの有無		合計	投与群	層 2: イベントの有無		合計
	あり	なし			あり	なし	
群 1	0	50	50	群 1	0	40	40
群 2	0	50	50	群 2	0	60	60
			100				100

次に、表 3.3 のように、全ての層において、「全ての群が 0%又は 100%」となった場合、FREQ プロシジャによって算出される Sato 信頼区間のみ出力され、他の信頼区間は出力されない。

表 3.3 Sato 信頼区間のみ出力される場合

投与群	層 1: イベントの有無		合計	投与群	層 2: イベントの有無		合計
	あり	なし			あり	なし	
群 1	0	50	50	群 1	40	0	40
群 2	50	0	50	群 2	0	60	60
			100				100

最後に、表 3.4 のように、いずれかの群において、「全ての層が 0%又は 100%」となった場合、FREQ プロシジャによって算出される Newcombe 信頼区間は出力されない。

表 3.4 Newcombe 信頼区間が出力されない場合

投与群	層 1: イベントの有無		合計	投与群	層 2: イベントの有無		合計
	あり	なし			あり	なし	
群 1	0	50	50	群 1	40	0	40
群 2	20	30	50	群 2	30	30	60
			100				100

次項のシミュレーションでは、いずれかの群で「全ての層が 0%」又は「全ての層が 100%」となった場合を除外して検討を行った。

4. シミュレーションの設定

本項では、Greenland 信頼区間、Sato 信頼区間、Newcombe 信頼区間のそれぞれの被覆確率をシミュレーションにより算出し、性能評価を試みる。層別項目の層の数は 2 層、4 層、8 層の場合に分けて、各群の例数やイベント発生割合について以下のようなシミュレーションを設定した。なお、信頼区間を算出するための信頼係数は 95%、シミュレーション回数は 10000 回とした。

2層の場合のシミュレーション (表 5.1 : No. 5.1.1.1~5.1.7.9)

- 例数については大きく分けて以下の2パターンを設定した。
 - ◇ 等例数 : 各群において各層の例数が等しい場合として、20例、50例、100例とした場合
 - ◇ 不等例数 : 各群において (20例、10例)、(50例、25例)、(100例、50例) とした場合
- 割合については大きく分けて以下の2パターンを設定した。
 - ◇ 等割合 : 各群において各層の割合が等しい場合として、10%、30%、50%とした場合
 - ◇ 不等割合 : 各群において、一方の層を50%、他方の層を5%、20%、40%とした場合

2層の場合のシミュレーション (表 5.1 : No. 5.1.8.1~5.1.8.14)

- 例数については大きく分けて以下の2パターンを設定した。
 - ◇ 等例数 : 各群において各層の例数が等しい場合として、10例とした場合
 - ◇ 不等例数 : 各群において (10例、50例) とした場合
- 割合については大きく分けて以下の2パターンを設定した。
 - ◇ 等割合 : 各群において各層の割合が等しい場合として、10%、50%とした場合
 - ◇ 不等割合 : 各群において、一方の層を50%、他方の層を10%とした場合

4層の場合のシミュレーション (表 5.2 : No. 5.2.1.1~5.2.7.9)

- 例数については大きく分けて以下の2パターンを設定した。
 - ◇ 等例数 : 各群において各層の例数が等しい場合として、20例、50例、100例とした場合
 - ◇ 不等例数 : 各群において3つの層を20例、50例、100例、残りの層をその半分の例数とした場合
- 割合については大きく分けて以下の2パターンを設定した。
 - ◇ 等割合 : 各群において各層の割合が等しい場合として、10%、30%、50%とした場合
 - ◇ 不等割合 : 各群において、3つの層を50%、残りの層を5%、20%、40%とした場合

4層の場合のシミュレーション (表 5.2 : No. 5.2.8.1~5.2.8.18)

- 例数については大きく分けて以下の2パターンを設定した。
 - ◇ 等例数 : 各群において各層の例数が等しい場合として、10例とした場合
 - ◇ 不等例数 : 各群において半分の層を20例、残りの半分の層を10例とした場合
- 割合については大きく分けて以下の2パターンを設定した。
 - ◇ 等割合 : 各群において各層の割合が等しい場合として、10%、50%とした場合
 - ◇ 不等割合 : 各群において、半分の層を10%、残りの半分の層を50%とした場合

8層の場合のシミュレーション (表 5.3 : No. 5.3.1~5.3.18)

- 例数については大きく分けて以下の2パターンを設定した。
 - ◇ 等例数 : 各群において各層の例数が等しい場合として、10例、20例とした場合
 - ◇ 不等例数 : 各群において半分の層を10例、残りの半分の層を20例とした場合
- 割合については大きく分けて以下の2パターンを設定した。
 - ◇ 等割合 : 各群において各層の割合が等しい場合として、10%、50%とした場合
 - ◇ 不等割合 : 各群において、半分の層を10%、残りの半分の層を50%とした場合

8層の場合のシミュレーション (表 5.3 : No. 5.3.19~5.3.22)

- 例数及び割合について、以下のパターンを設定した。
 - ◇ 不等例数：各群において半分の層を 10 例、残りの半分の層を 20 例、30 例、40 例とした場合
 - ◇ 不等割合：
 - ・ 各群において、半分の層を 10%、残りの半分の層を 40%、50%とした場合
 - ・ 各群において、半分の層を 5%、残りの半分の層を 50%とした場合

5. 被覆確率の比較

各シミュレーションの設定に基づき出力した結果を、層ごとに表 5.1 から 5.3 にまとめた。表には、Greenland 信頼区間、Sato 信頼区間、Newcombe 信頼区間のそれぞれにおける被覆確率を示している。また、各層の群間差が全て 0 である設定（真の群間差が 0 になる設定）においては、CMH 検定の結果「有意差なし」となった割合（ $1 - \alpha$ ）も併記した。なお、結果が 94.5%未満のものは格子掛け+太字 (**XX.XX**)、97.5%を超えたものは網掛け+下線 (XX.XX) にて表示した。

2層の場合のシミュレーション (表 5.1 : No. 5.1.1.1~5.1.7.9)

まず、Greenland 信頼区間においては、例数が小さい場合に被覆確率が 94.5%未満になることが多い傾向がみられたが、Newcombe 信頼区間では例数の影響はなかった。Sato 信頼区間においても、例数が小さい場合に被覆確率が 94.5%未満になることがあったが、Greenland 信頼区間よりも傾向は弱かった。

また、Sato 信頼区間は層内の割合の差が大きい場合 (No. 5.1.3.1、5.1.3.4、5.1.3.7 等) に被覆確率が 97.5%を超える傾向がみられた。Newcombe 信頼区間では、層間の割合の差が大きい場合 (No. 5.1.2.1、5.1.2.4、5.1.2.7、5.1.6.1、5.1.6.4、5.1.6.7 等) に被覆確率が 97.5%を超える傾向がみられた。

2層の場合のシミュレーション (表 5.1 : No. 5.1.8.1~5.1.8.14)

No. 5.1.1.1~5.1.7.9 のシミュレーション結果より、Greenland 信頼区間や Sato 信頼区間において例数が小さい場合に被覆確率が 94.5%未満になる傾向が示唆され、Sato 信頼区間や Newcombe 信頼区間においては割合の差が大きい場合に被覆確率が 97.5%を超える場合があったため、追加のシミュレーション No. 5.1.8.1~5.1.8.14 を行った。まず、Greenland 信頼区間においては、例数が小さい場合に被覆確率が 94.5%未満になることが多い傾向が明確になった。Sato 信頼区間も Greenland 信頼区間と同様の傾向がみられた一方、Newcombe 信頼区間では例数が小さい場合でも被覆確率が 94.5%未満になることは少なかった。

また、層内の割合の差が大きい場合のうち、No. 5.1.8.4、5.1.8.9 及び 5.1.8.14 等では、Sato 信頼区間にて被覆確率が大きくなる傾向がみられ、層間の割合の差が大きい場合のうち、No. 5.1.8.3、5.1.8.8 及び 5.1.8.12 等では Newcombe 信頼区間にて被覆確率が大きくなる傾向がみられたが、No. 5.1.8.13 の様に全ての信頼区間について被覆確率が 94.5%未満となる場合もみられることから、例数と割合のパターンにより結果が変わりうる点に注意する必要があることが分かる。

4層の場合のシミュレーション (表 5.2 : No. 5.2.1.1~5.2.7.9)

4層の設定のシミュレーションの結果では、各群の各層にある程度の例数がある場合、割合の設定に関わらず、3種類の信頼区間のいずれにおいても被覆確率は概ね 94.5%~97.5%の範囲内であったが、例数が少ない場合に Greenland 信頼区間のみ被覆確率が 94.5%未満となるが多かった。

4層の場合のシミュレーション (表 5.2 : No. 5.2.8.1~5.2.8.18)

シミュレーション No. 5.2.8.1~5.2.8.18 では各群の各層で例数が小さい場合に着目した。まず、No. 5.2.1.1~5.2.7.9の結果と同様、例数が小さい場合に Greenland 信頼区間は被覆確率が 94.5%未満となることが多かった。次に、Sato 信頼区間は、2層の際と同様、層内の割合の差が大きい場合 (No. 5.2.8.4, 5.2.8.9, 5.2.8.11 等) に被覆確率が 97.5%を超える傾向があった。また、Newcombe 信頼区間は、2層の際と同様、層間の割合の差が大きい場合に被覆確率が大きくなる傾向があることに加え、各群の各層の割合が小さい場合 (No. 5.2.8.1, 5.2.8.5, 5.2.8.12 等) に被覆確率が 97.5%を超える傾向があった。なお、シミュレーション No. 5.2.8.2 のように、各群の各層で例数が小さく、割合が大きい場合には、どの信頼区間も 94.5%未満の被覆確率となった。ちなみに、CMH 検定の結果も 94.5%を下回った。

8層の場合のシミュレーション (表 5.3 : No. 5.3.1~5.3.22)

最後に、8層の設定のシミュレーションでは、網羅的にシミュレーションを実施すると場合の数が膨大になるため、2層と4層の場合のシミュレーション結果を踏まえ、例数が小さい場合に着目して検討を行った。まず、Greenland 信頼区間においては、2層や4層の検討結果と同様、各群の各層の例数が小さい場合は、ほとんどのシミュレーション設定において被覆確率が 94.5%未満となったが、Sato 信頼区間及び Newcombe 信頼区間の被覆確率は、概ね 94.5%~97.5%の範囲内であった。また、Sato 信頼区間は、層内の割合の差が大きい場合 (No. 5.3.4, 5.3.8, 5.3.13 等) に被覆確率が 97.5%を超える傾向があった。なお、層内の割合の差が縮まれば、被覆確率は名義水準 (95%) に近づくことが No.5.3.19 より示唆される。さらに、No. 5.3.20 や 5.3.21 からは例数を増やすことにより、被覆確率が名義水準に近づく傾向であることが確認できる。Newcombe 信頼区間においては、2層や4層の検討結果と同様、層間の割合の差が大きい場合 (No. 5.3.18, 5.3.22 等) に被覆確率が大きくなる傾向があった。

シミュレーション結果のまとめ

各群の各層の例数が小さい場合には、Greenland 信頼区間の被覆確率は名義水準を下回る可能性が高くなる一方、Newcombe 信頼区間は層の数や例数によらず被覆確率を名義水準以上に維持することが示唆された。Rothman, Lash, and Greenland (2012b) では、各層の各群の分母が全て 2 以上であれば、Greenland 信頼区間を用いることが可能だと述べているが、シミュレーションの結果からは必ずしもそうではなく、例数が少ない場合においては Greenland 信頼区間の適用は注意が必要であるといえる。また、Sato 信頼区間は、Greenland 信頼区間と比べて、層の例数が小さいことによる影響を受けにくいことが示唆されたが、層内の割合の差が大きい場合、名義水準を上回る傾向も見受けられた。さらに、Newcombe 信頼区間においても、層間の割合の差が大きい場合には、名義水準を上回る傾向であった。

ちなみに、本項のシミュレーションは SAS マクロ %MH() にて実施した (SAS プログラムは本稿の最後で紹介)。前項のシミュレーションで設定した状況以外の場合においても、本マクロを使用することで 3 種類の信頼区間の被覆確率を計算することが出来る。実際の臨床試験計画時等で活用いただければ幸いである。

表 5.1 2層の場合の各設定における信頼区間の被覆確率

	No.	設定*		被覆確率			1- α
		例数	割合	Greenland	Sato	Newcombe	CMH 検定
等例数、等割合	5.1.1.1	(20, 20) (20, 20)	(10%, 10%) (10%, 10%)	95.67	95.77	<u>98.36</u>	96.62
	5.1.1.2	(20, 20) (20, 20)	(30%, 30%) (30%, 30%)	94.06	94.45	95.40	95.31
	5.1.1.3	(20, 20) (20, 20)	(50%, 50%) (50%, 50%)	94.60	94.64	94.64	94.64
	5.1.1.4	(50, 50) (50, 50)	(10%, 10%) (10%, 10%)	94.82	94.85	95.78	94.91
	5.1.1.5	(50, 50) (50, 50)	(30%, 30%) (30%, 30%)	94.80	94.89	95.09	95.19
	5.1.1.6	(50, 50) (50, 50)	(50%, 50%) (50%, 50%)	94.55	94.55	94.55	94.55
	5.1.1.7	(100, 100) (100, 100)	(10%, 10%) (10%, 10%)	95.03	95.12	95.49	95.46
	5.1.1.8	(100, 100) (100, 100)	(30%, 30%) (30%, 30%)	95.02	95.06	95.20	95.25
	5.1.1.9	(100, 100) (100, 100)	(50%, 50%) (50%, 50%)	95.20	95.20	95.20	95.20
等例数、層内で等割合、層間で不等割合	5.1.2.1	(20, 20) (20, 20)	(50%, 5%) (50%, 5%)	93.95	94.74	<u>98.17</u>	95.07
	5.1.2.2	(20, 20) (20, 20)	(50%, 20%) (50%, 20%)	94.18	94.55	96.49	95.46
	5.1.2.3	(20, 20) (20, 20)	(50%, 40%) (50%, 40%)	94.61	94.78	94.87	94.86
	5.1.2.4	(50, 50) (50, 50)	(50%, 5%) (50%, 5%)	94.52	94.83	<u>97.91</u>	95.08
	5.1.2.5	(50, 50) (50, 50)	(50%, 20%) (50%, 20%)	94.87	94.96	96.43	95.17
	5.1.2.6	(50, 50) (50, 50)	(50%, 40%) (50%, 40%)	94.93	94.94	94.95	94.95
	5.1.2.7	(100, 100) (100, 100)	(50%, 5%) (50%, 5%)	94.88	94.99	<u>97.77</u>	95.04
	5.1.2.8	(100, 100) (100, 100)	(50%, 20%) (50%, 20%)	94.93	95.01	96.28	95.21
	5.1.2.9	(100, 100) (100, 100)	(50%, 40%) (50%, 40%)	95.34	95.38	95.43	95.42
等例数、層内及び層間で不等割合	5.1.3.1	(20, 20) (20, 20)	(50%, 5%) (5%, 50%)	93.92	<u>98.34</u>	<u>97.95</u>	-
	5.1.3.2	(20, 20) (20, 20)	(50%, 20%) (20%, 50%)	93.91	96.57	96.47	-
	5.1.3.3	(20, 20) (20, 20)	(50%, 40%) (40%, 50%)	94.70	94.94	94.93	-
	5.1.3.4	(50, 50) (50, 50)	(50%, 5%) (5%, 50%)	94.59	<u>98.67</u>	<u>97.68</u>	-
	5.1.3.5	(50, 50) (50, 50)	(50%, 20%) (20%, 50%)	94.82	96.84	96.38	-
	5.1.3.6	(50, 50) (50, 50)	(50%, 40%) (40%, 50%)	94.79	94.87	94.80	-
	5.1.3.7	(100, 100) (100, 100)	(50%, 5%) (5%, 50%)	94.88	<u>98.83</u>	<u>97.82</u>	-
	5.1.3.8	(100, 100) (100, 100)	(50%, 20%) (20%, 50%)	94.94	96.98	96.18	-
	5.1.3.9	(100, 100) (100, 100)	(50%, 40%) (40%, 50%)	95.36	95.46	95.45	-

※ 各行の上段：(n₁₁, n₁₂) (p₁₁, p₁₂)、各行の下段：(n₂₁, n₂₂) (p₂₁, p₂₂)、 α ：Type I error rate

表 5.1 2層の場合の各設定における信頼区間の被覆確率 (続き)

	No.	設定*		被覆確率			1- α
		例数	割合	Greenland	Sato	Newcombe	CMH 検定
不等例数、等割合	5.1.4.1	(20, 10) (20, 10)	(10%, 10%) (10%, 10%)	97.19	97.26	<u>98.78</u>	97.33
	5.1.4.2	(20, 10) (20, 10)	(30%, 30%) (30%, 30%)	93.34	93.99	95.24	95.14
	5.1.4.3	(20, 10) (20, 10)	(50%, 50%) (50%, 50%)	94.57	94.98	95.15	95.12
	5.1.4.4	(50, 25) (50, 25)	(10%, 10%) (10%, 10%)	94.80	94.93	95.95	95.28
	5.1.4.5	(50, 25) (50, 25)	(30%, 30%) (30%, 30%)	94.60	94.84	95.34	95.31
	5.1.4.6	(50, 25) (50, 25)	(50%, 50%) (50%, 50%)	94.20	94.26	94.20	95.36
	5.1.4.7	(100, 50) (100, 50)	(10%, 10%) (10%, 10%)	94.95	95.01	95.66	95.15
	5.1.4.8	(100, 50) (100, 50)	(30%, 30%) (30%, 30%)	94.96	95.02	95.15	95.17
	5.1.4.9	(100, 50) (100, 50)	(50%, 50%) (50%, 50%)	94.41	94.45	94.41	94.97
不等例数、層内で等割合、層間で不等割合	5.1.5.1	(20, 10) (20, 10)	(50%, 5%) (50%, 5%)	93.77	95.04	<u>97.77</u>	95.52
	5.1.5.2	(20, 10) (20, 10)	(50%, 20%) (50%, 20%)	93.69	94.18	96.20	95.42
	5.1.5.3	(20, 10) (20, 10)	(50%, 40%) (50%, 40%)	94.28	94.85	95.21	95.17
	5.1.5.4	(50, 25) (50, 25)	(50%, 5%) (50%, 5%)	94.63	94.95	97.24	95.32
	5.1.5.5	(50, 25) (50, 25)	(50%, 20%) (50%, 20%)	94.69	94.91	95.71	95.27
	5.1.5.6	(50, 25) (50, 25)	(50%, 40%) (50%, 40%)	94.45	94.50	94.45	94.98
	5.1.5.7	(100, 50) (100, 50)	(50%, 5%) (50%, 5%)	94.86	94.99	97.47	95.07
	5.1.5.8	(100, 50) (100, 50)	(50%, 20%) (50%, 20%)	94.72	94.82	95.87	95.06
	5.1.5.9	(100, 50) (100, 50)	(50%, 40%) (50%, 40%)	94.67	94.70	94.67	94.80
不等例数、層内で等割合、層間で不等割合	5.1.6.1	(10, 20) (10, 20)	(50%, 5%) (50%, 5%)	93.01	95.03	<u>98.56</u>	95.37
	5.1.6.2	(10, 20) (10, 20)	(50%, 20%) (50%, 20%)	93.80	94.40	96.30	95.35
	5.1.6.3	(10, 20) (10, 20)	(50%, 40%) (50%, 40%)	94.13	94.78	95.27	95.21
	5.1.6.4	(25, 50) (25, 50)	(50%, 5%) (50%, 5%)	94.20	94.97	<u>98.18</u>	94.94
	5.1.6.5	(25, 50) (25, 50)	(50%, 20%) (50%, 20%)	94.54	94.73	96.15	95.16
	5.1.6.6	(25, 50) (25, 50)	(50%, 40%) (50%, 40%)	94.42	94.43	94.43	94.77
	5.1.6.7	(50, 100) (50, 100)	(50%, 5%) (50%, 5%)	94.75	95.10	<u>97.99</u>	95.12
	5.1.6.8	(50, 100) (50, 100)	(50%, 20%) (50%, 20%)	95.03	95.09	96.15	95.25
	5.1.6.9	(50, 100) (50, 100)	(50%, 40%) (50%, 40%)	94.85	94.85	94.85	94.86

※ 各行の上段：(n₁₁, n₁₂) (p₁₁, p₁₂)、各行の下段：(n₂₁, n₂₂) (p₂₁, p₂₂)、 α : Type I error rate

表 5.1 2層の場合の各設定における信頼区間の被覆確率 (続き)

	No.	設定*		被覆確率			1- α
		例数	割合	Greenland	Sato	Newcombe	CMH 検定
不等例数、層内及び層間で不等割合	5.1.7.1	(20, 10) (20, 10)	(50%, 5%) (5%, 50%)	93.36	<u>98.04</u>	<u>97.99</u>	-
	5.1.7.2	(20, 10) (20, 10)	(50%, 20%) (20%, 50%)	93.54	96.14	96.36	-
	5.1.7.3	(20, 10) (20, 10)	(50%, 40%) (40%, 50%)	94.29	95.11	95.28	-
	5.1.7.4	(50, 25) (50, 25)	(50%, 5%) (5%, 50%)	94.37	<u>98.38</u>	<u>97.72</u>	-
	5.1.7.5	(50, 25) (50, 25)	(50%, 20%) (20%, 50%)	94.67	96.62	96.12	-
	5.1.7.6	(50, 25) (50, 25)	(50%, 40%) (40%, 50%)	94.93	95.35	95.44	-
	5.1.7.7	(100, 50) (100, 50)	(50%, 5%) (5%, 50%)	94.80	<u>98.56</u>	<u>97.52</u>	-
	5.1.7.8	(100, 50) (100, 50)	(50%, 20%) (20%, 50%)	94.70	96.80	96.13	-
	5.1.7.9	(100, 50) (100, 50)	(50%, 40%) (40%, 50%)	94.65	94.82	94.65	-
追加検討	5.1.8.1	(10, 10) (10, 10)	(10%, 10%) (10%, 10%)	<u>98.98</u>	<u>99.12</u>	<u>99.78</u>	-
	5.1.8.2	(10, 10) (10, 10)	(50%, 50%) (50%, 50%)	92.29	93.16	96.00	-
	5.1.8.3	(10, 10) (10, 10)	(50%, 10%) (50%, 10%)	92.77	94.15	97.31	-
	5.1.8.4	(10, 10) (10, 10)	(50%, 10%) (10%, 50%)	92.83	97.39	97.43	-
	5.1.8.5	(50, 10) (50, 10)	(10%, 10%) (10%, 10%)	94.37	94.57	96.94	-
	5.1.8.6	(50, 10) (50, 10)	(50%, 50%) (50%, 50%)	94.48	94.50	94.50	-
	5.1.8.7	(50, 10) (50, 10)	(50%, 10%) (50%, 10%)	94.38	94.62	95.96	-
	5.1.8.8	(50, 10) (50, 10)	(10%, 50%) (10%, 50%)	94.22	94.85	96.89	-
	5.1.8.9	(50, 10) (50, 10)	(50%, 10%) (10%, 50%)	94.42	97.02	96.50	-
	5.1.8.10	(50, 10) (10, 50)	(10%, 10%) (10%, 10%)	94.75	95.12	<u>98.37</u>	-
	5.1.8.11	(50, 10) (10, 50)	(50%, 50%) (50%, 50%)	92.66	93.43	94.58	-
	5.1.8.12	(50, 10) (10, 50)	(50%, 10%) (50%, 10%)	92.22	93.12	97.30	-
	5.1.8.13	(50, 10) (10, 50)	(10%, 50%) (50%, 10%)	92.34	94.41	94.26	-
	5.1.8.14	(50, 10) (10, 50)	(50%, 10%) (10%, 50%)	94.48	<u>99.37</u>	<u>99.41</u>	-

※ 各行の上段 : (n_{11}, n_{12}) (p_{11}, p_{12}) 、各行の下段 : (n_{21}, n_{22}) (p_{21}, p_{22}) 、 α : Type I error rate

表 5.2 4層の場合の各設定における信頼区間の被覆確率

	No.	設定*		被覆確率			1- α
		例数	割合	Greenland	Sato	Newcombe	CMH 検定
等例数、等割合	5.2.1.1	(20,20,20,20)	(10%,10%,10%,10%)	94.24	94.89	96.71	95.10
		(20,20,20,20)	(10%,10%,10%,10%)				
	5.2.1.2	(20,20,20,20)	(30%,30%,30%,30%)	94.11	94.53	95.00	94.97
		(20,20,20,20)	(30%,30%,30%,30%)				
	5.2.1.3	(20,20,20,20)	(50%,50%,50%,50%)	94.13	94.81	94.92	94.91
		(20,20,20,20)	(50%,50%,50%,50%)				
	5.2.1.4	(50,50,50,50)	(10%,10%,10%,10%)	94.65	94.78	95.04	94.97
		(50,50,50,50)	(10%,10%,10%,10%)				
	5.2.1.5	(50,50,50,50)	(30%,30%,30%,30%)	94.54	94.69	94.78	94.85
(50,50,50,50)		(30%,30%,30%,30%)					
5.2.1.6	(50,50,50,50)	(50%,50%,50%,50%)	94.83	94.85	94.85	94.85	
	(50,50,50,50)	(50%,50%,50%,50%)					
5.2.1.7	(100,100,100,100)	(10%,10%,10%,10%)	94.58	94.75	94.83	94.81	
	(100,100,100,100)	(10%,10%,10%,10%)					
5.2.1.8	(100,100,100,100)	(30%,30%,30%,30%)	94.87	94.95	94.96	94.98	
	(100,100,100,100)	(30%,30%,30%,30%)					
5.2.1.9	(100,100,100,100)	(50%,50%,50%,50%)	94.70	94.70	94.70	94.70	
	(100,100,100,100)	(50%,50%,50%,50%)					
等例数、層内で等割合、層間で不等割合	5.2.2.1	(20,20,20,20)	(50%,50%,50%,5%)	94.50	94.99	97.11	95.49
		(20,20,20,20)	(50%,50%,50%,5%)				
	5.2.2.2	(20,20,20,20)	(50%,50%,50%,20%)	94.28	94.54	96.05	94.90
		(20,20,20,20)	(50%,50%,50%,20%)				
	5.2.2.3	(20,20,20,20)	(50%,50%,50%,40%)	93.98	94.87	95.09	95.06
		(20,20,20,20)	(50%,50%,50%,40%)				
	5.2.2.4	(50,50,50,50)	(50%,50%,50%,5%)	95.13	95.29	96.80	95.33
		(50,50,50,50)	(50%,50%,50%,5%)				
	5.2.2.5	(50,50,50,50)	(50%,50%,50%,20%)	94.81	94.88	95.93	95.02
(50,50,50,50)		(50%,50%,50%,20%)					
5.2.2.6	(50,50,50,50)	(50%,50%,50%,40%)	94.90	94.94	94.95	94.95	
	(50,50,50,50)	(50%,50%,50%,40%)					
5.2.2.7	(100,100,100,100)	(50%,50%,50%,5%)	94.96	95.05	96.84	95.17	
	(100,100,100,100)	(50%,50%,50%,5%)					
5.2.2.8	(100,100,100,100)	(50%,50%,50%,20%)	95.01	95.05	95.95	95.06	
	(100,100,100,100)	(50%,50%,50%,20%)					
5.2.2.9	(100,100,100,100)	(50%,50%,50%,40%)	94.89	94.89	94.89	94.89	
	(100,100,100,100)	(50%,50%,50%,40%)					
等例数、層内及び層間で不等割合	5.2.3.1	(20,20,20,20)	(50%,50%,50%,5%)	94.10	96.86	96.77	-
		(20,20,20,20)	(5%,50%,50%,50%)				
	5.2.3.2	(20,20,20,20)	(50%,50%,50%,20%)	94.13	95.83	96.06	-
		(20,20,20,20)	(20%,50%,50%,50%)				
	5.2.3.3	(20,20,20,20)	(50%,50%,50%,40%)	93.94	94.84	95.01	-
		(20,20,20,20)	(40%,50%,50%,50%)				
	5.2.3.4	(50,50,50,50)	(50%,50%,50%,5%)	94.85	96.91	96.76	-
		(50,50,50,50)	(5%,50%,50%,50%)				
	5.2.3.5	(50,50,50,50)	(50%,50%,50%,20%)	94.67	95.80	95.77	-
(50,50,50,50)		(20%,50%,50%,50%)					
5.2.3.6	(50,50,50,50)	(50%,50%,50%,40%)	94.81	94.90	94.90	-	
	(50,50,50,50)	(40%,50%,50%,50%)					
5.2.3.7	(100,100,100,100)	(50%,50%,50%,5%)	94.70	96.98	96.65	-	
	(100,100,100,100)	(5%,50%,50%,50%)					
5.2.3.8	(100,100,100,100)	(50%,50%,50%,20%)	94.74	95.72	95.67	-	
	(100,100,100,100)	(20%,50%,50%,50%)					
5.2.3.9	(100,100,100,100)	(50%,50%,50%,40%)	94.87	94.87	94.87	-	
	(100,100,100,100)	(40%,50%,50%,50%)					

※ 各行の上段：(n₁₁, n₁₂, n₁₃, n₁₄) (p₁₁, p₁₂, p₁₃, p₁₄)、各行の下段：(n₂₁, n₂₂, n₂₃, n₂₄) (p₂₁, p₂₂, p₂₃, p₂₄)、 α : Type I error rate

表 5.2 4層の場合の各設定における信頼区間の被覆確率 (続き)

	No.	設定*		被覆確率			1- α
		例数	割合	Greenland	Sato	Newcombe	CMH 検定
不等例数、等割合	5.2.4.1	(20,20,20,10) (20,20,20,10)	(10%,10%,10%,10%) (10%,10%,10%,10%)	94.27	95.11	97.54	95.70
	5.2.4.2	(20,20,20,10) (20,20,20,10)	(30%,30%,30%,30%) (30%,30%,30%,30%)	94.39	94.73	94.94	95.09
	5.2.4.3	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,50%) (50%,50%,50%,50%)	94.59	94.93	94.97	94.97
	5.2.4.4	(50,50,50,25) (50,50,50,25)	(10%,10%,10%,10%) (10%,10%,10%,10%)	94.78	94.92	95.47	95.15
	5.2.4.5	(50,50,50,25) (50,50,50,25)	(30%,30%,30%,30%) (30%,30%,30%,30%)	94.78	94.91	95.03	95.11
	5.2.4.6	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,50%) (50%,50%,50%,50%)	94.94	95.21	95.27	95.27
	5.2.4.7	(100,100,100,50) (100,100,100,50)	(10%,10%,10%,10%) (10%,10%,10%,10%)	94.91	95.02	95.20	95.08
	5.2.4.8	(100,100,100,50) (100,100,100,50)	(30%,30%,30%,30%) (30%,30%,30%,30%)	94.91	95.00	95.08	95.10
	5.2.4.9	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,50%) (50%,50%,50%,50%)	94.67	94.69	94.67	94.68
不等例数、層内で等割合、層間で不等割合	5.2.4.1	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,5%) (50%,50%,50%,5%)	94.46	94.69	96.57	95.12
	5.2.5.2	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,20%) (50%,50%,50%,20%)	94.26	94.98	95.56	95.44
	5.2.5.3	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,40%) (50%,50%,50%,40%)	94.58	94.95	95.03	95.02
	5.2.5.4	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,5%) (50%,50%,50%,5%)	94.84	95.21	96.54	95.54
	5.2.5.5	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,20%) (50%,50%,50%,20%)	94.74	94.91	95.82	95.10
	5.2.5.6	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,40%) (50%,50%,50%,40%)	94.77	95.20	95.32	95.29
	5.2.5.7	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,5%) (50%,50%,50%,5%)	95.34	95.40	96.15	95.41
	5.2.5.8	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,20%) (50%,50%,50%,20%)	94.97	95.16	95.34	95.23
	5.2.5.9	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,40%) (50%,50%,50%,40%)	94.79	94.79	94.79	94.79
不等例数、層内で等割合、層間で不等割合	5.2.6.1	(10,20,20,20) (10,20,20,20)	(50%,50%,50%,5%) (50%,50%,50%,5%)	94.28	95.07	97.14	95.58
	5.2.6.2	(10,20,20,20) (10,20,20,20)	(50%,50%,50%,20%) (50%,50%,50%,20%)	94.15	94.69	96.06	95.16
	5.2.6.3	(10,20,20,20) (10,20,20,20)	(50%,50%,50%,40%) (50%,50%,50%,40%)	94.27	94.73	94.80	94.80
	5.2.6.4	(25,50,50,50) (25,50,50,50)	(50%,50%,50%,5%) (50%,50%,50%,5%)	94.74	95.10	96.89	95.31
	5.2.6.5	(25,50,50,50) (25,50,50,50)	(50%,50%,50%,20%) (50%,50%,50%,20%)	94.69	94.89	95.87	94.99
	5.2.6.6	(25,50,50,50) (25,50,50,50)	(50%,50%,50%,40%) (50%,50%,50%,40%)	94.43	94.91	95.19	95.12
	5.2.6.7	(50,100,100,100) (50,100,100,100)	(50%,50%,50%,5%) (50%,50%,50%,5%)	95.04	95.13	96.99	95.19
	5.2.6.8	(50,100,100,100) (50,100,100,100)	(50%,50%,50%,20%) (50%,50%,50%,20%)	94.84	94.88	95.75	94.95
	5.2.6.9	(50,100,100,100) (50,100,100,100)	(50%,50%,50%,40%) (50%,50%,50%,40%)	94.74	94.74	94.74	94.74

※ 各行の上段：(n₁₁, n₁₂, n₁₃, n₁₄) (p₁₁, p₁₂, p₁₃, p₁₄)、各行の下段：(n₂₁, n₂₂, n₂₃, n₂₄) (p₂₁, p₂₂, p₂₃, p₂₄)、 α : Type I error rate

表 5.2 4層の場合の各設定における信頼区間の被覆確率 (続き)

	No.	設定*		被覆確率			1- α
		例数	割合	Greenland	Sato	Newcombe	CMH 検定
不等例数、層内及び層間で不等割合	5.2.7.1	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,5%) (5%,50%,50%,50%)	94.10	96.34	96.23	-
	5.2.7.2	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,20%) (20%,50%,50%,50%)	94.14	95.57	95.71	-
	5.2.7.3	(20,20,20,10) (20,20,20,10)	(50%,50%,50%,40%) (40%,50%,50%,50%)	94.50	95.02	95.05	-
	5.2.7.4	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,5%) (5%,50%,50%,50%)	94.67	96.66	96.55	-
	5.2.7.5	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,20%) (20%,50%,50%,50%)	94.73	95.70	95.62	-
	5.2.7.6	(50,50,50,25) (50,50,50,25)	(50%,50%,50%,40%) (40%,50%,50%,50%)	94.81	94.83	94.81	-
	5.2.7.7	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,5%) (5%,50%,50%,50%)	94.71	96.64	96.37	-
	5.2.7.8	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,20%) (20%,50%,50%,50%)	94.85	95.77	95.60	-
	5.2.7.9	(100,100,100,50) (100,100,100,50)	(50%,50%,50%,40%) (40%,50%,50%,50%)	94.77	94.82	94.77	-
追加検討	5.2.8.1	(10,10,10,10) (10,10,10,10)	(10%,10%,10%,10%) (10%,10%,10%,10%)	95.45	95.88	<u>98.96</u>	96.64
	5.2.8.2	(10,10,10,10) (10,10,10,10)	(50%,50%,50%,50%) (50%,50%,50%,50%)	93.66	94.28	94.30	94.30
	5.2.8.3	(10,10,10,10) (10,10,10,10)	(50%,50%,10%,10%) (50%,50%,10%,10%)	93.55	94.86	97.48	95.38
	5.2.8.4	(10,10,10,10) (10,10,10,10)	(50%,50%,10%,10%) (10%,10%,50%,50%)	93.08	<u>97.69</u>	97.37	-
	5.2.8.5	(20,20,10,10) (20,20,10,10)	(10%,10%,10%,10%) (10%,10%,10%,10%)	94.38	94.84	<u>97.82</u>	95.59
	5.2.8.6	(20,20,10,10) (20,20,10,10)	(50%,50%,50%,50%) (50%,50%,50%,50%)	94.35	94.64	94.61	94.61
	5.2.8.7	(20,20,10,10) (20,20,10,10)	(50%,50%,10%,10%) (50%,50%,10%,10%)	94.39	94.85	96.99	95.20
	5.2.8.8	(20,20,10,10) (20,20,10,10)	(10%,10%,50%,50%) (10%,10%,50%,50%)	93.78	94.65	97.30	95.03
	5.2.8.9	(20,20,10,10) (20,20,10,10)	(50%,50%,10%,10%) (10%,10%,50%,50%)	94.07	<u>97.72</u>	97.02	-
	5.2.8.10	(20,20,10,10) (20,20,10,10)	(50%,10%,50%,10%) (50%,10%,50%,10%)	94.11	95.01	97.42	95.41
	5.2.8.11	(20,20,10,10) (20,20,10,10)	(50%,10%,50%,10%) (10%,50%,10%,50%)	93.76	<u>97.88</u>	97.08	-
	5.2.8.12	(20,20,10,10) (10,10,20,20)	(10%,10%,10%,10%) (10%,10%,10%,10%)	93.84	94.87	<u>98.09</u>	95.67
	5.2.8.13	(20,20,10,10) (10,10,20,20)	(50%,50%,50%,50%) (50%,50%,50%,50%)	93.66	94.30	95.26	95.20
	5.2.8.14	(20,20,10,10) (10,10,20,20)	(50%,50%,10%,10%) (50%,50%,10%,10%)	93.67	94.57	97.39	95.17
	5.2.8.15	(20,20,10,10) (10,10,20,20)	(50%,50%,10%,10%) (10%,10%,50%,50%)	93.73	<u>98.83</u>	<u>98.58</u>	-
	5.2.8.16	(20,20,10,10) (10,10,20,20)	(10%,10%,50%,50%) (50%,50%,10%,10%)	93.19	96.49	95.73	-
	5.2.8.17	(20,20,10,10) (10,10,20,20)	(50%,10%,50%,10%) (50%,10%,50%,10%)	93.61	94.59	97.41	95.29
	5.2.8.18	(20,20,10,10) (10,10,20,20)	(50%,10%,50%,10%) (10%,50%,10%,50%)	93.37	<u>97.84</u>	97.21	-

表 5.3 8層の場合の各設定における信頼区間の被覆確率

No.	設定*		被覆確率			1- α
	例数	割合	Greenland	Sato	Newcombe	CMH 検定
5.3.1	(10,...,10, 10,...,10)	(10%,...,10%, 10%,...,10%)	94.05	95.55	<u>98.16</u>	95.90
	(10,...,10, 10,...,10)	(10%,...,10%, 10%,...,10%)				
5.3.2	(10,...,10, 10,...,10)	(50%,...,50%, 50%,...,50%)	92.93	94.90	95.25	95.23
	(10,...,10, 10,...,10)	(50%,...,50%, 50%,...,50%)				
5.3.3	(10,...,10, 10,...,10)	(50%,...,50%, 10%,...,10%)	93.44	94.73	97.29	95.21
	(10,...,10, 10,...,10)	(50%,...,50%, 10%,...,10%)				
5.3.4	(10,...,10, 10,...,10)	(50%,...,50%, 10%,...,10%)	93.84	<u>98.27</u>	97.41	-
	(10,...,10, 10,...,10)	(10%,...,10%, 50%,...,50%)				
5.3.5	(20,...,20, 20,...,20)	(10%,...,10%, 10%,...,10%)	94.49	95.06	95.94	95.20
	(20,...,20, 20,...,20)	(10%,...,10%, 10%,...,10%)				
5.3.6	(20,...,20, 20,...,20)	(50%,...,50%, 50%,...,50%)	94.48	95.11	95.11	95.11
	(20,...,20, 20,...,20)	(50%,...,50%, 50%,...,50%)				
5.3.7	(20,...,20, 20,...,20)	(50%,...,50%, 10%,...,10%)	94.67	95.18	97.17	95.31
	(20,...,20, 20,...,20)	(50%,...,50%, 10%,...,10%)				
5.3.8	(20,...,20, 20,...,20)	(50%,...,50%, 10%,...,10%)	94.62	<u>98.36</u>	97.47	-
	(20,...,20, 20,...,20)	(10%,...,10%, 50%,...,50%)				
5.3.9	(20,...,20, 10,...,10)	(10%,...,10%, 10%,...,10%)	94.17	94.91	96.60	95.31
	(20,...,20, 10,...,10)	(10%,...,10%, 10%,...,10%)				
5.3.10	(20,...,20, 10,...,10)	(50%,...,50%, 50%,...,50%)	94.05	94.99	95.66	95.51
	(20,...,20, 10,...,10)	(50%,...,50%, 50%,...,50%)				
5.3.11	(20,...,20, 10,...,10)	(50%,...,50%, 10%,...,10%)	94.31	95.01	96.62	95.14
	(20,...,20, 10,...,10)	(50%,...,50%, 10%,...,10%)				
5.3.12	(20,...,20, 10,...,10)	(10%,...,10%, 50%,...,50%)	93.69	94.85	97.42	95.10
	(20,...,20, 10,...,10)	(10%,...,10%, 50%,...,50%)				
5.3.13	(20,...,20, 10,...,10)	(50%,...,50%, 10%,...,10%)	94.43	<u>98.08</u>	97.14	-
	(20,...,20, 10,...,10)	(10%,...,10%, 50%,...,50%)				
5.3.14	(20,...,20, 10,...,10)	(10%,...,10%, 10%,...,10%)	93.85	95.01	96.84	95.14
	(10,...,10, 20,...,20)	(10%,...,10%, 10%,...,10%)				
5.3.15	(20,...,20, 10,...,10)	(50%,...,50%, 50%,...,50%)	93.64	94.68	94.99	94.96
	(10,...,10, 20,...,20)	(50%,...,50%, 50%,...,50%)				
5.3.16	(20,...,20, 10,...,10)	(50%,...,50%, 10%,...,10%)	93.43	94.49	97.30	94.84
	(10,...,10, 20,...,20)	(50%,...,50%, 10%,...,10%)				
5.3.17	(20,...,20, 10,...,10)	(10%,...,10%, 50%,...,50%)	93.61	94.68	97.24	95.12
	(10,...,10, 20,...,20)	(10%,...,10%, 50%,...,50%)				
5.3.18	(20,...,20, 10,...,10)	(50%,...,50%, 10%,...,10%)	94.40	<u>99.08</u>	<u>98.62</u>	-
	(10,...,10, 20,...,20)	(10%,...,10%, 50%,...,50%)				
5.3.19	(20,...,20, 10,...,10)	(40%,...,40%, 10%,...,10%)	94.35	97.16	96.70	-
	(20,...,20, 10,...,10)	(10%,...,10%, 40%,...,40%)				
5.3.20	(30,...,30, 10,...,10)	(50%,...,50%, 10%,...,10%)	94.74	<u>97.96</u>	97.00	-
	(30,...,30, 10,...,10)	(10%,...,10%, 50%,...,50%)				
5.3.21	(40,...,40, 10,...,10)	(50%,...,50%, 10%,...,10%)	94.67	<u>97.60</u>	96.68	-
	(40,...,40, 10,...,10)	(10%,...,10%, 50%,...,50%)				
5.3.22	(20,...,20, 10,...,10)	(50%,...,50%, 5%,...,5%)	94.13	<u>98.84</u>	<u>97.93</u>	-
	(20,...,20, 10,...,10)	(5%,...,5%, 50%,...,50%)				

※ 各行の上段：(n₁₁, ..., n₁₄, n₁₅, ..., n₁₈) (p₁₁, ..., p₁₄, p₁₅, ..., p₁₈)、各行の下段：(n₂₁, ..., n₂₄, n₂₅, ..., n₂₈) (p₂₁, ..., p₂₄, p₂₅, ..., p₂₈)
 α : Type I error rate

6. まとめ

本稿では、調整済みリスク差に対する信頼区間の算出方法を3種類紹介した上で、それぞれの信頼区間の性能評価を行った。シミュレーション結果より、Greenland 信頼区間、Sato 信頼区間、Newcombe 信頼区間のうち、層の数や例数によらず被覆確率が名義水準を維持したのは Newcombe 信頼区間であった。各群の各層の例数が小さい場合、Greenland 信頼区間は被覆確率が名義水準を下回る可能性が高いため、注意が必要である。Sato 信頼区間や Newcombe 信頼区間は、Greenland 信頼区間ほど例数による影響を受けないが、層内や層間の割合の設定によっては、被覆確率が名義水準を上回ることがあるため、層内及び層間の割合の差を考慮し、適用する信頼区間を選択することが望ましい。

参考文献

- Greenland S. and Robins J. M. (1985), "Estimation of a common effect parameter from sparse follow-up data," *Biometrics*, 41(1): 55-68.
- Sato T. (1989), "On variance estimator for the Mantel-Haenszel risk difference. (Letter)," *Biometrics*; 45: 1323-1324.
- Yan X. and Su X. G. (2010), "Stratified Wilson and Newcombe Confidence Intervals for Multiple Binomial Proportions," *Statistics in Biopharmaceutical Research*; Vol. 2, No. 3: 329-335.
- Rothman K. J. (2012a), "Epidemiology: An Introduction, 2nd Ed.," Oxford University Press.
- Rothman K. J., Lash T. L., and Greenland S. (2012b), "Modern Epidemiology, 3rd Ed.," Lippincott Williams & Wilki.
- SAS/STAT(R) 14.1 User's Guide.
- 飯塚 政人, 浜田 知久馬 (2013), "2 群の割合の差における信頼区間の構成法の比較," SAS ユーザー総会.
- 飯塚 政人, 魚住 龍史, 浜田 知久馬 (2014), "FREQ プロシジャによる割合の差の信頼区間," SAS ユーザー総会.

参考：5章のシミュレーションを実行する SAS マクロ

```
* _N1 : 群1の各層の例数をスペース区切りで入力 ;
* _N2 : 群2の各層の例数をスペース区切りで入力 ;
* _P1 : 群1の各層の割合をスペース区切りで入力 ;
* _P2 : 群2の各層の割合をスペース区切りで入力 ;
* _ITE : シミュレーション回数 ;
%macro MH(_N1=" 10, 20, 30", _N2=" 10, 20, 30", _P1="0.1, 0.2, 0.3", _P2="0.1, 0.2, 0.3", _ITE=10000) ;
options nosource nonotes ;
ods listing close ;

data _NULL_ ;
STR=0 ; FLAG=1 ;
do while (FLAG=1) ;
if (scan(&_N1., STR+1, ",")="") then FLAG=0 ;
else STR=STR+1 ;
end ;
call symputx("_STR",STR) ;
run ;

data SAMPLE ;
W=0 ; TRUE_D=0 ;
array N1 {&_STR.} ; array N2 {&_STR.} ;
```

```

array P1 {&_STR.} ; array P2 {&_STR.} ;
array X1 {&_STR.} ; array X2 {&_STR.} ;
do I=1 to &_STR. ;
  N1[I]=input(scan(&_N1., I, ", "), best12.) ;
  N2[I]=input(scan(&_N2., I, ", "), best12.) ;
  P1[I]=input(scan(&_P1., I, ", "), best12.) ;
  P2[I]=input(scan(&_P2., I, ", "), best12.) ;
  W =W+(N1[I]*N2[I])/(N1[I]+N2[I]) ;
end ;

do I=1 to &_STR. ;
  TRUE_D=TRUE_D+(P1[I]-P2[I])*(N1[I]*N2[I])/(N1[I]+N2[I])/W ;
end ;
call symputx("_TRUE_D",TRUE_D) ;

J=1 ;
do while (J<=&_ITE.) ;
  do I=1 to &_STR. ;
    X1[I]=ranbin(777, N1[I], P1[I]) ;
    X2[I]=ranbin(777, N2[I], P2[I]) ;
  end ;
  if (sum(of X1[*]) ne 0 and sum(of X2[*]) ne 0 and
      sum(of X1[*]) ne sum(of N1[*]) and sum(of X2[*]) ne sum(of N2[*])) then do ;
    do I=1 to &_STR. ;
      NO=1 ; TREAT=1 ; X=X1[I] ; N=N1[I] ; output ;
      NO=1 ; TREAT=2 ; X=X2[I] ; N=N2[I] ; output ;
      NO=2 ; TREAT=1 ; X=1 ; N=X1[I] ; output ;
      NO=2 ; TREAT=1 ; X=2 ; N=N1[I]-X1[I] ; output ;
      NO=2 ; TREAT=2 ; X=1 ; N=X2[I] ; output ;
      NO=2 ; TREAT=2 ; X=2 ; N=N2[I]-X2[I] ; output ;
    end ;
    J=J+1 ;
  end ;
end ;
keep J NO TREAT I N X ;
run ;

ods output Effect=CI ;
proc sdrate data=SAMPLE
  method=mh
  stat =risk
  effect=diff
  plots=none ;
  where NO=1 ;
  by J ;
  population group=TREAT event=X total=N ;
  strata I / order=internal stats(cl=normal) effect ;
run;
ods output close ;

data CP_STDRATE ;
  set CI ;
  if (LOWERCL <= &_TRUE_D. <= UPPERCL) then CP_WALD+(100/&_ITE.) ;
  if (_N_=&_ITE.) ;
  keep CP_WALD ;
run ;

ods output CommonPdiff=CI CMH=CMH ;
proc freq data=SAMPLE ;
  where NO=2 ;

```

```

    by J ;
    table I*TREAT*X / riskdiff(common column=1 cl=newcombe) cmh nocol nopercnt plots=none ;
    weight N ;
run ;
ods output close ;

data CP_SATO ;
    set CI ;
    where (Method="Mantel-Haenszel") ;
    if (LOWERCL <= &_TRUE_D. <= UPPERCL) then CP_SATO+(100/&_ITE.) ;
    if (_N_=&_ITE.) ;
    keep CP_SATO ;
run ;

proc sql noprint ;
    select count(*) into:_ITE_NEW from CI where Method="Newcombe" ;
quit ;

data CP_NEW ;
    set CI ;
    where (Method="Newcombe") ;
    if (LOWERCL <= &_TRUE_D. <= UPPERCL) then CP_NEW+(100/&_ITE_NEW.) ;
    if (_N_=&_ITE_NEW.) ;
    keep CP_NEW ;
run ;

data CP_CMH ;
    set CMH ;
    where (Statistic=1) ;
    if (Prob>=0.05) then CP_CMH+(100/&_ITE.) ;
    if (_N_=&_ITE.) ;
    keep CP_CMH ;
run ;

options source notes ;
ods listing ;
data CP ;
    merge CP_STDRATE CP_SATO CP_NEW CP_CMH ;
run ;

title1 "N1 =" &_N1. " ; N2 =" &_N2. " ;
title2 "P1 =" &_P1. " ; P2 =" &_P2. " ;
title3 "True Difference=&_TRUE_D." ;
proc print ; run ;
    title1 "" ; title2 "" ; title3 "" ;
%mend ;

* 2層の指定例 ;
%MH(_N1=" 20, 20", _N2=" 20, 20",
    _P1="0.1, 0.1", _P2="0.1, 0.1") ;

* 4層の指定例 ;
%MH(_N1=" 20, 20, 20, 20", _N2=" 20, 20, 20, 20",
    _P1="0.1, 0.1, 0.1, 0.1", _P2="0.1, 0.1, 0.1, 0.1") ;

* 8層の指定例 ;
%MH(_N1=" 10, 10, 10, 10, 10, 10, 10, 10", _N2=" 10, 10, 10, 10, 10, 10, 10, 10",
    _P1="0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1", _P2="0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1") ;

```

以上