

ワクチン開発における非劣性試験のデザインについて

○江口凌平¹、浦狩保則¹

(¹シミック(株)データサイエンス事業本部)

Design of Non-Inferiority Clinical Trials in Vaccine Development

Ryohei Eguchi, Yasunori Uragari

Data Science Div., CMIC Co.,Ltd.

要旨

新規ワクチン群と対照薬群の非劣性試験の計画において、リスク比を治療効果の尺度とした場合の非劣性試験のデザインについて検討した。非劣性試験では、適切な非劣性マージンを担保した上で、試験薬のワクチン有効性を評価しなければならない。非劣性マージンを設定する上で、FDA ガイドライン(2016)で述べられている手法に従うことが最良であると考えた。リスク比についての非劣性マージンの設定のからくり(対数尺度での解釈、リスク比が望小特性であることなど)を図解した。

非劣性マージンの設定事例として、Senn(2022)による複数の過去の試験のメタアナリシスを行った。そこでは、恒常性(constancy assumption)を考慮して、変量効果モデルの当てはめを行い、非劣性マージンの導出を行った。算出した非劣性マージンをもとに、必要イベント数・必要症例数のシミュレーションを行い、イベント発現率が大きい場合、小さい場合に算出される必要症例数についても考察を行った。

キーワード：ワクチン試験、非劣性試験、非劣性マージン、必要イベント数、ワクチン有効性、リスク比

1. ワクチン試験について

パンデミック下では、新型コロナウイルスワクチンのプラセボ対照試験において、プラセボ投与群の被験者には実薬が投与されないため、被験者が最善の医療を得られるような倫理的な配慮が必要とされた。このようなプラセボ投与群を設定することが困難な場合に、すでにコロナワクチンとして承認されている実薬を比較対象として臨床試験を実施することが考えられる。そこで、すでに有効性が判明しているワクチンを対照薬とにおいて、免疫原性の指標、主に中和抗体価をエンドポイントとした、ランダム化実薬対照非劣性試験、又は優越性試験を代替とした開発が進められている。2021年6月24日に開催された国際薬事規制当局連携組織(International Coalition of Medicines Regulatory Authorities: ICMRA)のワークショップでの議論によれば、このような試験では、ワクチンの免疫原性とその発症予防効果との間の高い相関性を示す必要がある。そのためには、免疫原性(抗体価の上昇)がサロゲートマーカーとして使用可能であること(中和抗体価の測定方法の信頼性確保、実薬と新たなワクチンとの間の細胞性免疫の反応の違いの考慮など)、国際的な評価基準(EMA, FDA など)に基づいていること、および複数の臨床試験で一貫した結果が得られていることなどの条件が必要

である(原 他(2013), 渡邊 他(2018))。

パンデミック後の現在は、新型コロナウイルス感染症の位置づけが「5類感染症」になり、社会の向き合い方も大きく変化した。ワクチン試験の実施についても、発症予防効果をエンドポイントとした臨床試験の実施の動きがある。2群試験に限れば、プラセボ対照優越性試験か実薬対照非劣性試験の選択がある。なお、上記の ICMRA ワークショップでも、発症予防効果をエンドポイントとしたプラセボ対照試験が依然として重要であることが結論付けられている。ただし、冒頭で述べた倫理的問題を考慮すれば、発症予防効果をエンドポイントとする実薬対照非劣性試験の計画・実施の可能性は考えられる。

本稿では、実薬対照非劣性試験における必要な適切な非劣性マージンの設定、適切な必要イベント数の決定について検討する。

2. リスク比を評価指標とする場合の非劣性マージン設定の手順

一般に非劣性試験とは、試験薬 (T)が対照薬 (C)に比べて非劣性マージン (δ)以上に劣ることはないことを示す試験のことである。帰無仮説 H_0 と対立仮説 H_1 とすると、 $H_0: C-T \geq \delta$ 、 $H_1: C-T < \delta$ とそれぞれ表すことができる。ワクチン試験において、有効性の評価指標としてワクチン有効性が用いられる。ワクチン有効性は、ワクチンにおける発現率のプラセボにおける発現率に対するリスク比(Risk Ratio)を1から減じた値、すなわち1-リスク比で定義されている。今回はリスク比を治療効果の尺度として使用することを想定する。

ここでは、上で定義されるプラセボに対するリスク比ではなく、対照薬群に対するリスク比、すなわち(試験薬群のイベント発現率: P_T)/(実対照薬群のイベント発現率: P_C)を用いる。非劣性仮説は以下の通り示すことができる。(ここでの発現率は発現割合の意味合いで用いている。)

$$H_0: \frac{P_T}{P_C} \geq M \text{ vs } H_1: \frac{P_T}{P_C} < M \quad (1)$$

また、上式を対数変換すると、同様に以下の通り示すことができる。

$$H_0: \log P_T - \log P_C \geq \log M \text{ vs } H_1: \log P_T - \log P_C < \log M \quad (2)$$

非劣性試験では、非劣性マージン M の設定が重要である。非劣性マージンは、対照薬 vs. プラセボの既存の試験で観測された群間差をもとに、当該試験において、プラセボ以上の効果が間接的に示されるような値に設定すべきことが、FDA ガイドラインでも明確に述べられている。具体的には、過去に行われた対照薬 vs. プラセボの優越性試験の差を M_1 、 M_1 の何割が維持されていれば試験薬が受け入れ可能かの臨床的判断を反映するための M_2 という指標を用いて非劣性マージンを算出する。今回は固定マージン法による M_1 と M_2 の選定に着目する。

試験薬 vs. 対照薬における M_1 と M_2 の選定手順について、図1を用いて説明する。図1中に記載の括弧書きの数値については、後ほど説明する。試験薬 vs. 対照薬の試験を検討する上で、①対照薬 vs. プラセボ、②プラセボ vs. 対照薬、③試験薬 vs. 対照薬の3つの段階に分けて考える。

① 対照薬 vs. プラセボ: 対照薬群のイベント発現率を P_C 、プラセボ群のイベント発現率を P_p とすると、対照薬群のプラセボ群に対するリスク比は、 P_C/P_p となる。以降、リスク比が望小特性であることに注意する。

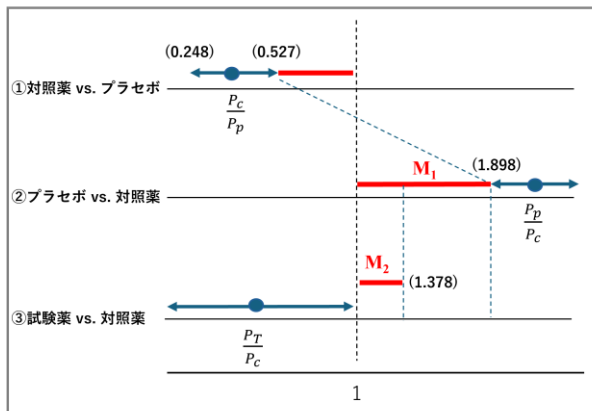
このリスク比をもとに非劣性マージンの算出を行うが、対数尺度上で解釈することが重要である。右図の①はリスク比を対数変換した $\log P_C - \log P_p$ を用いた場合の図である。

② プラセボ vs. 対照薬: ②は、プラセボ群と対照薬群の関係を①とは逆にしており、最後の③における試

試験薬 vs. 対照薬の前段階の図である。FDA ガイドラインでは、左図の②にある P_p/P_c の 95%信頼区間の下限値を非劣性マージン M_1 と置いている。右図に示すように、対数尺度上では、①での 0 から $\log P_c - \log P_p$ の 95%信頼区間の上限値までの距離と②での $\log P_p - \log P_c$ の 95%信頼区間の下限値から 0 までの距離が等しい、すなわち①と②の赤い線分の長さが等しいことに留意したい。

- ③ 試験薬 vs. 対照薬：③は、②でのプラセボ群を試験薬群に置き換えた形になっている。左図の③に示す P_T/P_c のリスク比について、その信頼区間の上限 $< M_1$ であることを、試験薬がプラセボよりも大きい効果を持つこと、つまり有効性の証明として位置づけることができる。ただし、臨床的判断のもとでは、その信頼区間の上限が M_1 よりも小さいだけでは十分ではなく、 $M_2 < M_1$ となるような M_2 を非劣性マージンとすることが必要であるとされる。通常、 M_2 を M_1 の 50%とすることが多いが、この計算は対数変換尺度上で行われていることに留意したい。実際に、右図の③で、 $\log M_2 = \frac{1}{2} \log M_1$ として M_2 の値が求められる。

原尺度上での解釈



対数尺度上での解釈

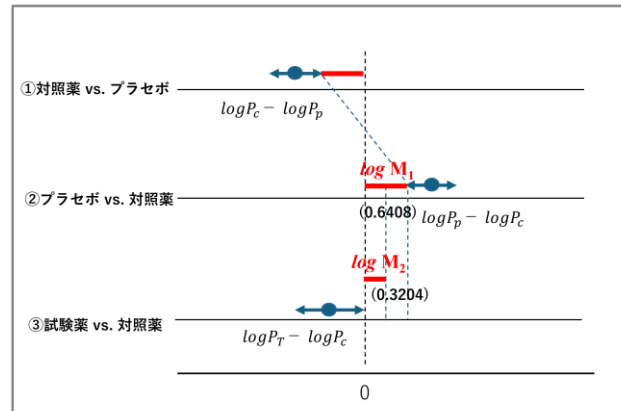


図1. リスク比を指標とする場合の非劣性マージン M_1 と M_2 の選定手順

M_1 と M_2 を算出する事例が、FDA ガイドライン(2016)で示されている。その事例では、新しい抗凝固剤 ximelagatran と対照薬ワーファリンを用いた非劣性試験(SPORTIF V)を取り上げている。まずは、固定マージン法を用いた非劣性マージン M_1 の算出を記述している。 M_1 を算出するために、変量効果メタアナリシスを用いて、6つのプラセボ対照試験のそれぞれのリスク比を統合し、リスク比の点推定値 0.361 とその 95%信頼区間(0.248, 0.527)を得ている。この変量効果メタアナリシスの手法については、次の3節で詳細に記載する。95%信頼区間の上限値 0.527 は、図1の左図の①に示す対照薬のプラセボに対するリスク比 P_c/P_p の 95%信頼区間の上限値に該当する。その逆数の $1/0.527 = 1.898$ は、左図の②に示すプラセボの対照薬に対するリスク比 P_p/P_c の 95%信頼区間の下限値に該当する。ガイドラインでは、もし試験薬とプラセボの効果が同じであれば何が起こるかという観点から、プラセボであれば対照薬に対して約 90%のリスクを増大させる(すなわち、リスク比が 1.898 となる)と説明している。つまり、左図の③で、試験薬の対照薬に対するリスク比 P_T/P_c が 1.898 よりも小さいことが、試験薬がプラセボよりも優れた効果をもつことを意味している。

続いて、その事例では M_2 の算出を記述している。具体的に、試験薬が最大限に許容できる臨床的マージン M_2 を M_1 の 50%とするとの設定のもとで、図1の右図の③の通り、 $\log M_1 = \log 1.898 = 0.6408$ の 50%は 0.3204 と算出される。これを原尺度上での値に変換すると、 $M_2 = \exp(0.3204) = 1.378$ となる。なお、SPORTIF V 試験では、試験薬の対照薬に対するリスク比の点推定値が 1.39 であり、その両側 95%信頼区間が (0.91, 2.12)であった。上限値(2.12)が $M_2(1.378)$ よりも大きいため、この事例において ximelagatran のワーファ

リンに対する非劣性は示されなかったと説明されている。続く3節では、2節で紹介した手順に従い、文献で公表されているデータを用いながら非劣性マージンの設定について紹介する。

3. 実データに基づく非劣性マージンの設定と試験の選択

非劣性試験のデザインでは、分析感度の検討が必須である。分析感度は治療間の差を識別する能力(ability)であり、仮に今の非劣性試験にプラセボ群を含めていたと想定すると、分析感度として、対照薬とプラセボの間に少なくとも M_1 の差が存在したはずであることを意味している。このような M_1 の選択や試験が分析感度を持っているかどうかを判断する上で、歴史的エビデンス(HESDE : Historical evidence of sensitivity to drug effects)を用いた変量効果メタアナリシスの手法が FDA ガイドラインでは推奨されている。HESDE は、過去に適切にデザインされ実施された試験で、特定の実治療を用いて、その治療がプラセボ群あるいは他の治療群よりも優れていることを示した試験を指す。過去の複数の試験で一貫して見られていることをもって、プラセボ群あるいは他の治療群に対してその治療が優れていると解釈する。ここで、上記の HESDE の情報を使用するためには、今計画している非劣性試験と過去の試験の類似性が担保されていなければならない。この仮定を、「恒常性の仮定(constancy assumption)」と呼ぶ。

我々は Senn (2022)の5試験のデータから変量効果メタアナリシスの手法を用いて、一つの包括的な統計量を算出する。この5試験は全て大規模なランダム化プラセボ対照試験であり、表1に過去5試験のイベント発現状況についてまとめている。

表1. 過去5試験のイベント発現状況

Study	Subjects		Cases		追跡期間
	Vaccine	Placebo	Vaccine (%)	Placebo (%)	
A	20,712	21,096	77 (0.37)	850 (4.03)	6 ヶ月
B	17,662	8550	73 (0.41)	130 (1.52)	2 ヶ月
C	14,134	14,073	11 (0.08)	185 (1.31)	2 ヶ月
D	7020	7019	10 (0.14)	96 (1.37)	3 ヶ月
E	19,630	19,691	116 (0.59)	348 (1.77)	2 ヶ月

変量効果メタアナリシスのモデルとして、以下に記載の Bakbergenuly et al.(2019)の二項一般化線形混合モデルを用いる。

$$\log(P_{ji}) = \alpha_j + (\theta + b_j)x_i, b_j \sim N(0, \tau^2) \quad j = 1, \dots, K; i = C, T \quad (3)$$

ここで、 j は 1~ K 番目の試験、 i は対照薬群(C)、試験薬群(T)をそれぞれ表し、 $x_c = 0$, $x_T = 1$ とする。そして、 P_{ji} は試験 j における群 i の発現率、 α_j は試験 j におけるプラセボ群の対数発現率を表す固定効果、および $(\theta + b_j)$ は試験 j におけるワクチン群のプラセボ群に対する対数リスク比である。ここで、 θ は全体の対数リスク比、 $b_j \sim N(0, \tau^2)$ は変量効果、および τ^2 は試験間分散を指している。したがって、このモデルのあてはめにおいて、 α_j 、 θ および τ^2 を推定する。上記のモデルに対応する SAS プログラムを以下に記載する。

```

proc glimmix data=targ;
  class study treatment;
  model event/n = study treatment / noint link=log solution cl;
  random treatment / subject=study;
  lsmeans treatment / diff=all cl ilink;
  covtest 'common variance' ZEROG;
  output out=out2 pred(noilink)=predMu resid(noilink)=residMu
  LCL(noilink)=lower UCL(noilink)=upper;
  ods output diffs=diffsall;
run;

```

上記のように、GLIMMIX プロシジャ(SAS Ver.9.4)を用いて、Bakbergenuly et al.(2019)の二項一般化線形混合モデルをあてはめた。以下にプログラムの要点を記載する。

- 非劣性マージンを設定する上で、恒常性の仮定を満たすことは重要であるため、プラセボ群に対する対照薬群の処理効果が試験間でどれほど変動するかに関心がある。そこで、処理効果の部分に変量効果を加え、random ステートメントには treatment を記載している。一方、プラセボ効果については、固定効果とした。(なお、Bakbergenuly et al.(2019)にはプラセボ効果も変量効果とするモデルも示されているが、2つの変量間の相関も考慮することが必要である。)
- この解析では、対数リスク比にてワクチン有効性を推定することから、link 関数を log に設定している。
- 対数リスク比が対数変換尺度上では引き算で求めることができるため、diff オプションを指定している。
- covtest 'common variance' ZEROG; にて、変量効果の分散=0 の仮定が妥当かどうか、すなわち恒常性を仮定することの妥当性を評価している。

さらに、以下の2点を補足しておく。

- 一般的にリスク比のメタアナリシスで二項一般化線形混合モデルを用いる場合、各試験のイベント発現率が区間(0,1)に存在するという自然な制約が守られず、試験間変動の分散や対数リスク比の推定においてバイアスが招来され得ること、そして対照群の発現率が中程度かつ/あるいは対数リスク比が大きな値である場合に、バイアスが問題であることが報告されている(Bakbergenuly et al.,2019)。しかし、ワクチン試験では、イベント発現率が表 1 でも示される通り、0.08%~0.59%とかなり小さいため、そのバイアスは起こらないとしても良い。(起こらないことを確認している。)
- 表 1 に示したように、各試験の追跡期間は一樣ではなかった。実際に、プラセボ群の発現率は、追跡期間が最長の試験 A が他の 4 試験よりも高値となっており、これらの試験の解析では追跡期間の長さも考慮することが必要であると考えられた。そのため、model ステートメントの右辺に、追跡期間を表す変数 followup を追加した場合も実行したが、変数 followup の回帰係数は 0 で推定された。その理由は、各試験のプラセボ効果を表す変数 study と変数 followup が別名(alias)の関係になっており、変数 study の効果の推定値に followup の効果が込みにして含まれているためであると考えた。

上記モデルを用いて、各試験および全体の対数リスク比とその 95%信頼区間を算出したところ、図 2 のような結果が得られた。

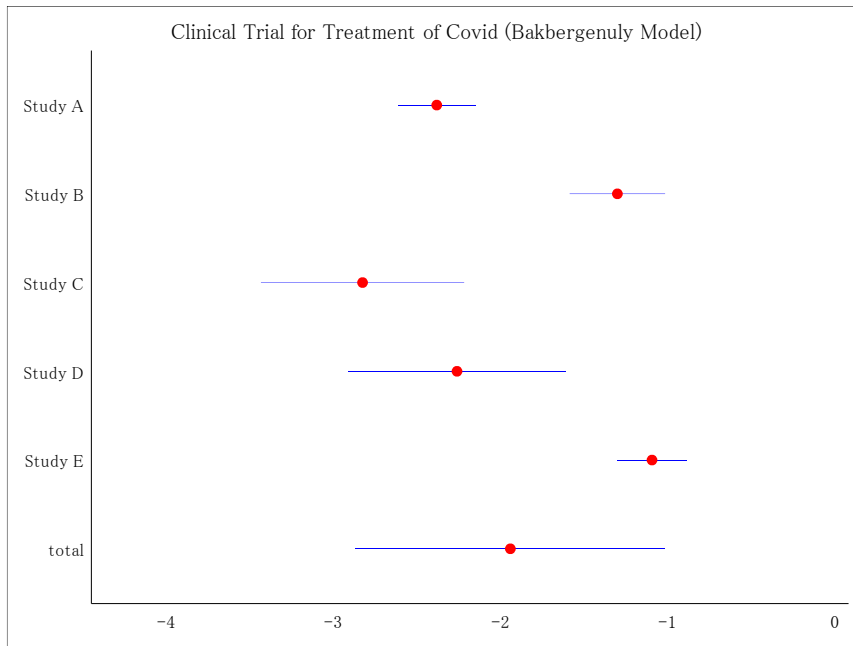


図 2. 過去 5 試験および全体の対数リスク比とその 95%信頼区間

Total がメタアナリシスを用いた全体の対数リスク比の値であり、 -1.9428 と算出された。

また、上図の結果を見ると、他試験と比べて、Study B と Study E は明らかに対数リスク比の値が大きいが分かった。これらの試験を含めた解析を行うことは、恒常性の仮定から外れた解析になっている可能性があるため、Study B と Study E は除いた解析も行った。

Study B と Study E は除いた上で、各試験および全体の対数リスク比とその 95%信頼区間を算出したところ、図 3 のような結果が得られた。

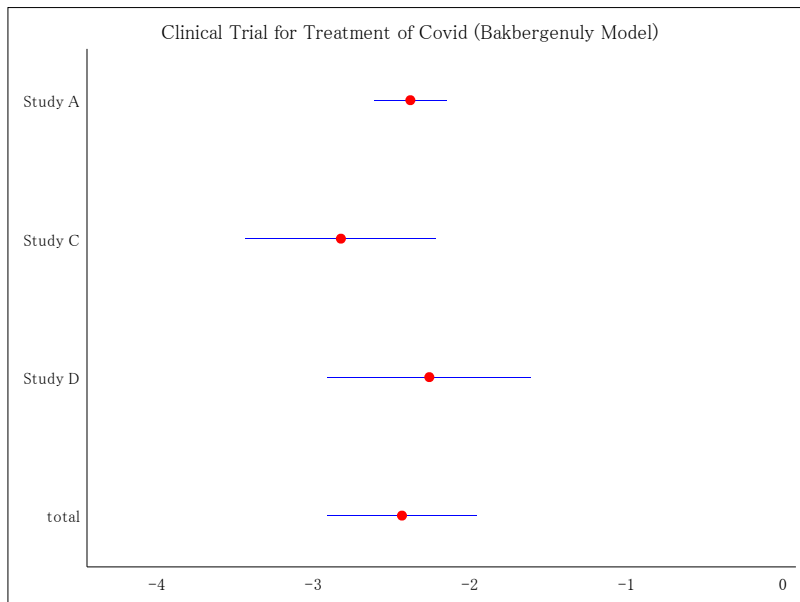


図 3. 過去 3 試験および全体の対数リスク比とその 95%信頼区間

ここで、先ほどの 5 試験全てを含めた解析と結果を比較したところ、表 2、表 3 の結果が得られた。

表 2. 対数リスク比の推定値

Study	5 試験全てを含む		3 試験のみを含む	
	Estimate	95%CI	Estimate	95%CI
A	-2.3831	(-2.6155, -2.1506)	-2.3831	(-2.6155, -2.1506)
B	-1.3026	(-1.5881, -1.0171)	-	-
C	-2.8268	(-3.4346, -2.2190)	-2.8268	(-3.4346, -2.2190)
D	-2.2619	(-2.9123, -1.6115)	-2.2619	(-2.9123, -1.6115)
E	-1.0955	(-1.3047, -0.8863)	-	-
Total	-1.9428	(-2.8695, -1.0160)	-2.4366	(-2.9144, -1.9588)

表 3. COVTEST 結果

5 試験全てを含む			3 試験のみを含む		
共分散パラメータ 推定値 τ^2	標準誤差	P-value	共分散パラメータ 推定値 τ^2	標準誤差	P-value
0.2561	0.195	<0.0001	0.001	0.027	0.4828

表 2 において推定値の結果を比較したところ、先ほどの 5 試験全てを含めた解析と比べ、全体の対数リスク比は小さくなり、-2.4366 と算出された。

また、COVTEST の結果を比較したところ、5 試験全てを含めた場合の COVTEST の検定結果は $p < 0.0001$ と有意な結果になっているのに対し、3 試験のみを含めた場合には、COVTEST の検定結果は $p = 0.4828$ となり、有意ではない結果となっている。この結果から、対数リスク比の値が大きい試験を除いた上での解析については、非劣性試験に必要である恒常性の仮定が満たされたと判断し、2 節で記述した考え方を元に、 M_1 と M_2 の非劣性マージンを算出する。

上述で求められた Total の推定値の信頼区間の上限値に該当する -1.9588 の値を用いる。-1.9588 は対照薬 vs. プラセボでの値であるため、2 節で述べた通り、プラセボ vs. 対照薬を考慮し、 $\log M_1 = 1.9588$ となる。臨床的にはその半分が維持されるべきであると考え、 $\log M_2 = (0.5)\log M_1 = 0.9794$ となる。原尺度へ戻すと、 $M_2 = \exp(0.9794) = 2.6629$ が求められる。すなわち、新たなワクチンを開発する上で、試験薬の対照薬に対するリスク比の信頼区間の上限値が、算出した非劣性マージン値の 2.6629 よりも低い場合に非劣性が証明されることになる。

ところで、Fleming et al.(2021)は、類似の方法を用いて、非劣性マージンを算出している。具体的に、対照薬のプラセボに対するワクチン有効性が 0.9、すなわちリスク比が 0.1 の場合(これは、表 1 の試験 A および試験 C におけるワクチン有効性に対応する)、そのプラセボ対照試験でのイベント数が 175 例および 350 例であった場合について、新規ワクチンの対照薬に対する非劣性マージンをそれぞれ 2.561 および 2.724 とすることが示されている。これらの値は、我々が算出した非劣性マージンに酷似している。また、Fleming et al.(2021)は非劣性マージンを 3 よりも大きくすべきでないとのステークホルダの見方を考慮して、データ解析で得られる非劣性マージンを M として、試験デザインに用いる非劣性マージンを $\min(M, 3)$ で与えることを提示している。この考え方は、ICH-E10 (2001)の「非劣性試験における限界値の決定は、統計的推論と臨床判断の双方

に基づくものであり、その選択の基となった証拠の不確実性を反映するとともに、適度に保守的であるべきである。」に基づいている。

4. 条件付き法による必要イベント数・必要被験者数の決定

3 節にて算出した非劣性マージンを用いて、必要イベント数・必要被験者数を決定した。必要イベント数は、浦狩(2025)で述べられている方法を元にして算出した。すなわち、ワクチン群とプラセボ群の 2 群比較試験の計画(割付け比は 1 対 1)において、2 群のイベント発現数合計に条件付きでワクチン群のイベント発現数が従う二項分布に基づいて、ワクチン有効性についての仮説検定で必要なイベント発現数を算出する方法(条件付き法)を基本とした。その際、2 群の割付け比を 1 対 1、片側有意水準を 0.025、検出力を 0.9 とした。優越性試験では、試験ワクチンの有効性を π として、検定の帰無仮説と対立仮説を

$$H_0: \pi \leq \pi_0 \text{ vs } H_1: \pi > \pi_0 \quad (4)$$

とし、帰無仮説値 π_0 、対立仮説の期待値 $\pi_1 (> \pi_0)$ を指定すれば必要イベント数(2 群合計)を算出できる。

本稿で対象としている非劣性試験では、検定の帰無仮説と対立仮説が

$$H_0: RR' \geq r_0 \text{ vs } H_1: RR' < r_0 \quad (5)$$

で表される。ここで、 RR' は試験ワクチンの対照ワクチンに対するリスク比、 r_0 は非劣性マージン((1)式のMと同一)であり、これを帰無仮説値とおく。2 つのワクチンの有効性が等しいとの仮定のもとで対立仮説の期待値を $\pi_1=1$ とおくのが一般的である。この検定仮説を、(プラセボではなく)対照ワクチンを基準とするワクチン有効性 $\pi' = 1 - RR'$ を用いて書き換えると、上記の非劣性検定の帰無仮説と対立仮説は

$$H_0: \pi' \leq \pi'_0 \text{ vs } H_1: \pi' > \pi'_0 \quad (6)$$

で表される。ここで、帰無仮説値は $\pi'_0 = 1 - r_0$ であり、対立仮説の期待値を $\pi'_1 = 1 - r_1 = 0$ とおく。これらの設定のもとで条件付き法を用いれば、非劣性試験での必要イベント数を算出することができる。なお、非劣性マージンが $r_0 > 1$ であるため、帰無仮説値は $\pi'_0 < 0$ となるが対応するケース・プロポーション θ (総発現イベント数に条件付きの二項分布のパラメータ)について $0 < \theta_0 = (1 - \pi'_0)/(2 - \pi'_0) < 1$ である。また、対立仮説の期待値 π'_1 に対応するケース・プロポーションは $\theta_1 = 0.5$ である。したがって、帰無仮説と対立仮説のそれぞれについて、条件付き二項分布の確率計算が問題なく実行可能である。詳細は、浦狩(2025)を参照。

3 節で算出した非劣性マージン $r_0 = M_2 = 2.6629$ から $\pi'_0 = 1 - r_0 = -1.6629$ および $\pi'_1 = 0$ とし、片側有意水準 0.025 で 0.9 以上の検出力を得るために必要な総イベント数(T_{req})が 53 イベント、棄却限界値(Y_c : 非劣性が認められる試験ワクチン群のイベント数の最大値)が 31 イベントと算出される(表 4)。なお、Fleming et al.(2021)は、非劣性マージンが 2.561 および 2.724 に対応する必要な総イベント数がそれぞれ 54 例および 48 例であることを示している。したがって、表 4 での結果は Fleming et al.(2021)で示された結果と同等のものが示されたことが分かった。

表 4. 非劣性試験で必要な総イベント数(2 群合計)

ワクチン有効性による記述		必要総イベント数 : T_{req}	棄却限界値 : Y_c
π'_0	π'_1		
-1.6629	0	53	31

必要被験者数の算出は、2 つのワクチンのもとのイベント発現率 Λ_v (件/人年)が同じであることを仮定し、浦狩(2025)の(8)式のプラセボの発現率 Λ_p を Λ_v で置き換えることで行える。

ところで、Fleming et al.(2021)には、ワクチンの非劣性試験はプラセボ対照試験よりも試験規模が 2~3 倍に

なることを示す次の記載があり、荒木(2021)、厚生労働省(2021)でも引用されている。

“... the 34-event trial comparing two vaccines having approximate 95% vaccine efficacy and the 48-event trial comparing two vaccines having approximate 90% efficacy would require approximately two to three-fold person years of follow-up relative to a frequently used design of a 150-event placebo-controlled trial of a vaccine having 60% vaccine efficacy, assuming these trials were conducted in settings having similar attack rates.”

ここでは、上記で算出した非劣性試験での必要イベント数 53 に対応する被験者数の算出の傍らで、この記載内容を定量的に確認することとした。その際、以下の条件設定を行った。

- ・ プラセボ群のイベント発現率(件/人年) : 0.01、0.05、0.1。ここでの 0.1(件/人年)はパンデミック時の発現率に対応している(表 1 を参照)。
- ・ ワクチン群のイベント発現率(件/人年) : プラセボ群のイベント発現率の 10 分の 1。3 節で算出した対照ワクチンの対プラセボのリスク比の推定値 $0.0875(=\exp(-2.4366))$ であることに基づく。この設定は、“two vaccines having approximate 90% efficacy” の状況に近いものである。
- ・ プラセボ対照優越性試験：“150-event placebo-controlled trial of a vaccine having 60% vaccine efficacy” については、(4)式において、対立仮説の期待値を $\pi_1=0.6$ として、必要イベント数が 150 となる帰無仮説値 π_0 を算出した。この条件を満たす π_0 は 0.285(必要イベント数は 150)で算出された(優越性試験①)。別に、 $\pi_1=0.6$ 、 $\pi_0=0.3$ とおくと必要イベント数は 160 になる(優越性試験②)。これら 2 つの優越性試験①、②を検討対象とした。
- ・ 非劣性試験 : 3 節で算出した非劣性マージンのもとでの非劣性試験の必要イベント数 53 の非劣性試験を “the 48-event trial comparing two vaccines having approximate 90% efficacy” に近い試験であると位置づけた。
- ・ 試験期間 : 0.5 年、1 年の 2 通りで必要被験者数を算出した。

上記の条件設定のもと優越性試験と非劣性試験における必要被験者数を算出し、結果を表 5 に示す。

表 5. 優越性試験と非劣性試験の必要被験者数

プラセボ群の発現率(件/人年)	ワクチン群の発現率(件/人年)		必要イベント数	必要被験者数(2 群合計)	
				試験期間 0.5 年	試験期間 1 年
0.01	0.001	優越性試験①	150	54,546	27,274
		優越性試験②	160	58,182	29,092
		非劣性試験	53	106,000	53,000
0.05	0.005	優越性試験①	150	10,910	5,456
		優越性試験②	160	11,638	5,820
		非劣性試験	53	21,200	10,600
0.1	0.01	優越性試験①	150	5,456	2,728
		優越性試験②	160	5,820	2,910
		非劣性試験	53	10,600	5,300

優越性試験① : 帰無仮説値 $\pi_0=0.285$ 、対立仮説の期待値 $\pi_1=0.6$

優越性試験② : 帰無仮説値 $\pi_0=0.3$ 、対立仮説の期待値 $\pi_1=0.6$

非劣性試験 : 非劣性マージン $r_0 = M_2=2.6629$ 、帰無仮説値 $\pi'_0 = -1.6629$ 、対立仮説の期待値 $\pi'_1 = 0$

表 5 の結果が示すように、試験期間が 0.5 年(1 年)の場合、非劣性試験では 53 イベントを集積するために、プラセボ群の発現率(件/人年)が 0.01、0.05 および 0.1 のとき、必要な被験者数はそれぞれ 106,000 例(53,000 例)、

21,200 例(10,600 例)および 10,600 例(5,300 例)と算出された。これらの必要被験者数は、優越性試験のおよそ 2 倍であることが読み取れる。すなわち、ワクチン有効性が高い(例えば 90%)状況では、非劣性試験ではプラセボよりもイベント発現率が顕著に低い(例えば、プラセボの 10 分の 1)2 つのワクチンを用いるため、必要イベント数に到達するまでに多くの観察人年が必要であると考察した。

5. まとめ

本稿では、新規ワクチン群と対照薬群の非劣性試験の計画において、リスク比を治療効果の評価指標とした場合の非劣性試験のデザインについて検討した。

具体的には、恒常性の仮定を考慮し、文献で公表されている複数の過去試験のメタアナリシスにおいて、変量効果モデルの当てはめを行い、非劣性マージンを算出した。本稿では、変量効果モデルとして、GLIMMIX プロシジャ(SAS Ver.9.4)を用いて、Bakbergenuly et al.(2019)の二項一般化線形混合モデルで最も単純なモデルを適用した。Bakbergenuly et al.(2019) においては、他にもいくつかの変量効果モデルが提案されているが、モデルが複雑になるにつれパラメータ数が増加し、推定が困難になる場合が考えられる。そのため、今回はパラメータ数が少なく、最も単純な二項一般化線形混合モデルを適用した。

また、上記で算出した非劣性マージンの設定を行う上で、我々は、①対照薬 vs. プラセボ、②プラセボ vs. 対照薬、③試験薬 vs. 対照薬の 3 つの段階に分け、原尺度上、対数尺度上の双方での解釈を行った。FDA ガイドライン(2016)においては、数式を用いない文言のみの表記で非劣性マージンの算出が説明されているため、解釈しづらい箇所が散見される。実際に非劣性マージンを導出する際は、本稿における図解の通り、算出された値がどの尺度であり、①、②、③のどの段階の値を指しているのかを理解しておく必要がある。また、算出した非劣性マージンを用いて、必要イベント数・必要被験者数を決定する際は、浦狩(2025)で述べられている条件付き法の手法を適用した。

ここでは、ワクチン開発における非劣性試験を取り上げたが、リスク比の尺度を用いた試験全般に本稿で示した手順は適用が可能である。FDA ガイドラインには、そのような薬剤として、抗凝固剤と急性心筋梗塞の治療薬が挙げられているが、他にも該当する薬剤が存在すると考えている。

参考文献

- Senn S.(2022). The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy. *Pharmaceutical Statistics*, 21,790-807.
- FDA Guidance Document (2016). Non-Inferiority Clinical Trials to Establish Effectiveness. <https://www.fda.gov/media/78504/download>
- Fleming TR, Krause PR, Nason M, Longini IM, Henao-Restrepo AM. (2021). COVID-19 vaccine trials: The use of active controls and non-inferiority studies. *Clin Trials*, 18(3): 335-342.
- Bakbergenuly I, Hoaglin DC, Kulinskaya E, et al. (2017). Pitfalls of using the risk ratio in meta-analysis. *Res Synth Methods*, 10(3): 398-419.
- 浦狩保則(2025). ワクチン試験での必要イベント数の設定について. SAS ユーザー総会.
- 原めぐみ, 大藤さとこ, 福島若葉, 廣田良夫(2013). ワクチン疫学研究の原理と方法: 新型インフルエンザ ワクチンの免疫原性と有効性の評価. 日本衛生学会
- 渡邊直昭, 大島信治, 永井英明, 加藤貴史, 齋藤美奈子, 五十嵐彩夏, 鈴川真穂, 川島正裕, 浅里功, 松井弘稔(2018). 成人における 4 価インフルエンザ不活化スプリットワクチンの免疫原性と安全性の検討, 日本呼吸器学会誌, 7 (4), 197-203.

- ICH-E10(2001). 臨床試験における対照群の選択とそれに関連する諸問題.
<https://www.pmda.go.jp/files/000156634.pdf>
- 荒木康弘(2021). これからの新型コロナウイルスワクチンの開発・評価について. 学術の動向, 第26巻第10号, 51-57.
- 厚生労働省(2021). ICMRA 活動と今後の開発・評価について. ワクチンに関する説明会資料(令和3年7月15日).