

別添 3

厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）
分担研究報告書

生物統計学的な観点からのワクチン開発における治験計画の立案の迅速化のための研究

分担課題：ワクチンの有効性評価におけるベイズ流統計手法の有用性に関する研究

研究分担者： 坂巻顕太郎

所属：横浜市立大学大学院データサイエンス研究科

研究要旨：

ベイズ流統計手法により、ワクチンの効果（VE: Vaccine Efficacy）の推測における不確実性を確率的に表現できる。これにより、「VE が 30%以上である可能性」などを数値化して表現できるようになり、中間解析における試験終了などをその可能性を考慮して判断できるようになる。そのため、p 値などと比べ、非統計家にとって試験結果の解釈が容易になるという利点がある。また、事前の規定とは異なる理由で試験が早期中止されるような場合でも、結果の不確実性を確率的に表現することで様々な判断を容易にする可能性がある。本研究では、海外で実施された SARS-CoV2 ワクチンの検証的試験での適用事例を踏まえ、国内のワクチン開発におけるベイズ流統計手法の利用可能性を検討した。

実際の検証的試験では、モニタリングやデータマネジメントなどの試験実施における実務的な側面での負荷を減らすようなベイズ流統計手法が使われていた。また、第一種の過誤確率が制御される手法であることを確認した上で適用されていた。SARS-CoV2 ワクチンの開発には使われていなかったが、ヒストリカルデータの利用に関する議論もあり、今後の適用が期待される。ワクチン開発にベイズ流統計手法を適用できるようにするためには、平常時の承認申請においても規制当局などがベイズ流統計手法を利用可能な状況にしておく準備が推奨される。これにより、新興感染症の流行下において国内でのワクチン開発において効率的な試験の実施が可能となる。

A. 研究目的

ワクチンの有効性を検証する試験では、対象とする感染症の特徴によって、検証的試験のサンプルサイズが大きくなる、試験期間が長くなるなどの問題が生じる可能性がある。実際、SARS-CoV-2 に対するワクチンの第3相臨床試験でランダム化された対象者数は、Pfizer/BioNTech の試験では43,548人[1]、Moderna の試験では30,420人[2]、AstraZeneca の試験では32,451人[3]、Johnson & Johnson の試験では43,788人[4]であった。新興感染症が流行している状況下では有効なワクチンを迅速に承認することが重要であるため、効率的な試験デザインや解析方法が必要となる。その方法の一つとして、ベイズ流統計手法の利用が考えられる。

ベイズ流統計手法を用いることの利点はいくつかあるが、例えば、中間解析などのプロトコルで規定した解析に対応するだけでなく、感染状況、治療やワクチンの開発状況に応じて必要となる早期の解析などで生じる解析結果の不確実性を考慮した解釈ができることがある。その他にも、過去に実施された試験のデータ（ヒストリカルデータ）の活用などもできる。

本研究では、SARS-CoV-2 に対するワクチンの有効性を評価する検証的試験で用いられたベイズ流統計手法の統計的な背景をレビューし、国内のワクチン開発においてベイズ流統計手法を利用する利点と課題について整理する。

B. 研究方法

SARS-CoV-2 に対するワクチンの一つである BNT162b2 (Pfizer/BioNTech) の臨床

試験は、第1/2/3相試験という形式で実施された。第3相試験における有効性の評価では、SARS-CoV-2 を感染した証拠のない対象での VE (Vaccine Efficacy) が 95.0% (95%信用区間：90.3% - 97.6%)、既感染者も含む対象での VE が 94.6% (95%信用区間：89.9% - 97.3%) という結果が示された[1]。

第3相試験では、VE の推測（有効性の有無の判断と効果の推定）に対してベータ二項分布を用いたベイズ流の推測を行っている (C. 研究結果)。また、最終解析に加えて3回の中間解析（プロトコルでは4回）を行う試験デザインを組んでおり、各時点での有効中止に対する判断はベイズ流統計手法に基づいて行っている。この試験デザインを用いるために、頻度論流に第一種の過誤確率が制御される方法であることを確認している。これは、FDA (U.S. Food and Drug Administration) をはじめとする規制当局が要請する、検証的試験（第3相試験）で用いる統計手法が持つべき特徴である。

本研究では、BNT162b2 の第3相試験のプロトコルをレビューし、プロトコルに記載のない統計的な背景を説明する。その上で、国内のワクチン開発においてベイズ流統計手法を用いる利点などをまとめる。

C. 研究結果

1. Vaccine Efficacy (VE) の定義

ワクチンの有効性を評価する指標の一つである VE は、

$$VE = 100 \times (1 - IRR)$$

と定義される。IRR は infection rate ratio (感染率比) や incidence rate ratio (発生

率比)の略で、プラセボ群での SARS-CoV-2 の発生率を IR_0 、ワクチン群での SARS-CoV-2 の発生率を IR_1 とすると、

$$IRR = IR_1 / IR_0$$

である。発生率は「単位時間当たりの発生数」として計算される。VE は $-\infty \sim 1$ の間の値をとり、0 の場合は効果がなく、1 (100%) に近づくほど効果が大きいことを意味する。

BNT162b2 の試験では、VE が 30%以上であるときにワクチンが有効であると判断する基準を置いている。

2. VE のベイズ流の評価

VE の推測における不確実性を考慮するためには、データから計算される VE (点推定値) だけで何らかの推測を行うのでは不十分である。ベイズ流統計手法を用いると、不確実性を確率的に考慮した推測が行える。例えば、「VE が 30%以上である可能性 (確率) は 99%である」というような推測が行える。正確には、この可能性は観察されたデータのもとで推測するため、

$$\Pr[VE \geq 30\% \mid \text{観察データ}]$$

という、条件付き確率を用いた推測を行うことになる。

ベイズ流統計手法では、推測したいパラメータ (ここでは VE という効果指標) に対して、事前に想定される確信度のようなもの (パラメータの事前分布) を用いる。例えば、VE がどのような値になるか事前にまったくわからなければ、VE が各値をとる可能性 (確率) は同等、VE が 30%であることを事前に完全に確信しているのであれば、VE = 30%である確率は 100%、などを事前に想定したことを踏まえた推測

を行う。事前分布を観察データと一緒に用いて VE の不確実性を表現する (パラメータの事後分布を計算する) のがベイズ流統計手法である。

BNT162b2 の試験では、ワクチン接種から SARS-CoV2 の発生までの時間 (time-to event) に対してモデルを仮定するのではなく、それぞれの治療群における SARS-CoV2 の発生数がポワソン分布という確率分布に従うことを仮定している。この場合、ワクチン群の発生数が二項分布という確率分布に従うと想定できることから、VE の推測を行う (別添資料 1)。さらに、

$$\theta = (1-VE)/(2-VE)$$

という変換から、 θ の事前分布として、ベータ分布 (Beta(0.700102, 1)) を用いた推測を行っている。 θ は 0 から 1 の値をとる。Beta(0.700102, 1) の 95%信用区間は (0.005, 0.964) であり、これは θ (または VE) が取りうる値の範囲のほとんどを覆っている。そのため、VE に対する事前の確信度はほとんどないと想定していたことになる。

実際の試験では、事前分布と観察されたデータから、VE が 30%を超える可能性を計算した結果、99.99%以上であることが示され、BNT162b2 は有効であると判断された[1]。

3. 中間解析と最終解析における判断

各解析時点では、VE に関する確信度がどの程度あるかによってワクチンが有効であるかを判断することになる。BNT162b2 の試験では、4 回の中間解析では VE が 30%以上である可能性が 99.50%を超える場合

$\Pr[VE \geq 30\% \mid \text{観察データ}] > 99.50\%$
に、最終解析では VE が 30%以上である可能性が 98.60%を超える場合

$\Pr[VE \geq 30\% \mid \text{観察データ}] > 98.60\%$
にワクチンが有効であると判断し、試験を中止（終了）することとしている。

BNT162b2 のデザインを用いる場合、各解析時点は試験全体での SARS-CoV2 発生数で決めることができる。この試験では、4 回の中間解析は、32 人、62 人、92 人、120 人で SARS-CoV2 の発生が確認された時点で、最終解析は 164 人で SARS-CoV2 の発生が確認された時点で行うこととしている。また、ワクチンが有効であると判断する基準は、前述の試験終了の定義から、ワクチン群での SARS-CoV2 の発生数から再定義できる。この試験では、それぞれの解析時点で、6 人以下、15 人以下、25 人以下、35 人以下、53 人以下であればワクチンが有効であると判断できる。この判断基準を用いると第一種の過誤確率が 2.5% 以下になる。このデザインではベータ二項分布を想定しているため、第一種の過誤確率は明示的に計算できる（別添資料 2）。

D. 考察

検証的試験においてベイズ流統計手法を適用することが日本ではいままでなかったため、本研究では、海外試験での適用からワクチン開発におけるベイズ流統計手法の利用可能性を検討した。BNT162b2 の試験では、ベイズ流統計手法を用いて VE の不確実性を確率的に表現し、ワクチンの有効性を判断した。この方法では、発生数のデータのみを用いて確率が計算できるため、SARS-CoV2 に感染するまでの時間などを

モニタリングする必要はない。そのため、モニタリングやデータマネジメントなどの試験実施における実務的な側面での負荷を減らす有用な方法になりうる。また、解析結果に基づく有効性の判断において、VE が 30%以上になる可能性などを数値化して表現できるため、非統計家にとって結果の解釈が容易になるという利点もベイズ流統計手法にはある。基礎的、臨床的、疫学的な知見が十分に集積されるまでは、試験実施に影響する要因などの情報が不確実な状況で試験をデザインすることが想定されるため、実際の試験では事前の規定とは異なる理由で試験が早期中止される可能性がある。このような場合でも、結果の確信度（不確実性）を確率的に表現できるため、ベイズ流統計手法は様々な判断を容易にする可能性がある。

BNT162b2 の試験で用いられた手法にはいくつか留意すべき点がある。その一つは効果指標の定義である。この試験ではポワソン分布の仮定から発生率比を用いて VE を計算していたが、Moderna [2]、AstraZeneca [3]、Johnson & Johnson [4] の試験では IRR はハザード比により規定されている。より柔軟に SARS-Cov2 の発生に対するモデルを考えるには、ハザード比を用いるほうが良い可能性があるため、その点を留意する必要がある。また、解析時点（中間解析の回数）、解析時点での有効性の判断基準の決定にも注意が必要である。この試験で用いた方法は、頻度論流の中間解析で用いるアルファ消費関数などに基づいていないため、試験デザインの事前規定や途中変更には統計的性質に基づいた評価が必要になる。BNT162b2 の試験では明示

的に計算可能なデザインを用いているが、デザインや効果指標によってはシミュレーションなどの評価が必要になることに留意しなければならない。

SARS-CoV2 のワクチン開発には使われていなかったが、ワクチンの開発に有用と考えられているその他のベイズ流統計手法として、ヒストリカルデータの利用がある[5, 6]。新興感染症が流行している状況下において、ヒストリカルデータをどう利用するか議論はまだ成熟していないが、リアルワールドデータ、リアルワールドエビデンスの利用に関する議論[7]などから、今後議論が進んでいくと考えられる。

E. 結語

ベイズ流統計手法を用いることで効率的に臨床試験を実施することが可能となる。特に、試験実施を容易にしたり、中間解析を含む意思決定に関する判断基準が明確になったりと、柔軟な臨床試験を組むことが可能となる。新興感染症の流行下では、不確実な情報により試験をデザインすることが想定され、規定外の早期中止をする可能性があるなど、通常の臨床試験実施体制とは異なる状況が想定される。そのため、不確実性がある状況で臨床的に有意義な結果が得られているかを判断するための強力な手段としてベイズ流統計手法の適用が考えられる。ワクチン開発にベイズ流統計手法を適用できるようにするためには、平常時の承認申請においても規制当局がベイズ流統計手法を利用可能な状況にしておく準備が推奨される。これにより、新興感染症の流行下において国内でワクチンを開発する際に効率的な試験の実施が可能となる。

F. 研究発表

特になし

G. 知的所有権の取得状況

特になし

<参考文献>

1. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine* 2020; 383(27): 2603-2615.
2. Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine* 2021; 384: 403-16.
3. Falsey AR, Sobieszczyk ME, Hirsch I, et al. Phase 3 safety and efficacy of AZD1222 (ChAdOx1 nCoV-19) Covid-19 vaccine. *New England Journal of Medicine* 2021; 385(25): 2348-2360.
4. Sadoff J, Gray G, Vandebosch A, et al. Final analysis of efficacy and safety of single-dose Ad26. COV2. S. *New England Journal of Medicine* 2022; 386(9): 847-860.
5. Liu M, Li Q, Lin J, et al. Innovative trial designs and analyses for vaccine clinical development. *Contemporary Clinical Trials* 2021; 100: 106225.
6. Jin M, Feng D, Liu G. Bayesian approaches on borrowing historical data for vaccine efficacy trials. *Statistics in Biopharmaceutical Research* 2020; 12(3): 284-292.

7. U.S. Food and Drug Administration. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. Draft guidance for industry. Available at <https://www.fda.gov/media/154714/> download. (Accessed 2022/4/18)

別添資料1 SARS-CoV-2の発生数と二項分布の関係

Vaccine Efficacy (VE) は

$$VE = 100 \times (1 - IRR)$$

と定義される。IRR は infection rate ratio (感染率比) や incidence rate ratio (発生率比) の略で、プラセボ群での SARS-CoV-2 の発生率を IR_0 、ワクチン群での SARS-CoV-2 の発生率を IR_1 とすると、

$$IRR = IR_1 / IR_0$$

である。ここで、 $VE = 1 - IRR$ とし、以下の変換を考える。

$$\begin{aligned}\theta &= (1-VE)/(2-VE) \\ &= (1-1+IRR)/(2-1+IRR) \\ &= IRR/(1+IRR) \\ &= (IR_1/IR_0)/(1+(IR_1/IR_0)) \\ &= IR_1/(IR_0+IR_1)\end{aligned}$$

ここで、プラセボ群での観察人年を T_0 、SARS-CoV-2 の発生数を E_0 、ワクチン群での観察人年を T_1 、SARS-CoV-2 の発生数を E_1 とすると、

$$(E_1/T_1)/(E_0/T_0 + E_1/T_1)$$

と $IR_1/(IR_0+IR_1)$ を推定できる。さらに、 E_0 、 E_1 がポワソン分布に従い、観察人年が等しいとすると、

$$\begin{aligned}(E_1/T_1)/(E_0/T_0 + E_1/T_1) \\ = E_1/(E_0 + E_1)\end{aligned}$$

となる。このとき、分母の十分統計量 ($E_0 + E_1$) を条件づけると、 E_1 が従う分布は二項分布になる。

別添資料2 BNT162b2 試験における第一種の過誤確率を計算するためのRプログラム

```
### On: March, 2022
### By: Kentaro Sakamaki
### For: 第一種の過誤確率の計算

### 関数の引数の説明 ###
#VE: VE
#total.event.X: X 回目の解析時点での観察イベント数
#crit.X: X 回目の解析時点で試験治療群が優れていると判断するイベント数の上限

### 第一種の過誤確率を計算する関数 ###
alpha <- function(VE,
  total.event.1, total.event.2, total.event.3, total.event.4, total.event.5,
  crit.1, crit.2, crit.3, crit.4, crit.5
){

  TE <- list()
  TE[[1]] <- total.event.1; TE[[2]] <- total.event.2; TE[[3]] <- total.event.3; TE[[4]] <-
total.event.4; TE[[5]] <- total.event.5;

  CV <- list()
  CV[[1]] <- crit.1; CV[[2]] <- crit.2; CV[[3]] <- crit.3; CV[[4]] <- crit.4; CV[[5]] <- crit.5;

  #2 項分布のパラメータ
  theta = (1-VE)/(2-VE)

  #各インターバルで観察されるイベント数
  N <- list()
  N[[1]] <- TE[[1]];
  for (i in 1:4) N[[i+1]] <- TE[[i+1]] - TE[[i]];

  #棄却確率を保存するためのリスト
  P <- list()
```


#1 回目の解析で生じる第一種の過誤確率

```
pb <- NULL;
for (event.group1 in 0:CV[[1]]) {
  pb <- c(pb, choose(N[[1]], event.group1)*(theta**event.group1)*((1-theta)**(TE[[1]]-
event.group1)))
}
P[[1]] <- pb
```

#2 回目以降の解析で生じる第一種の過誤確率

```
for (i in 1:4){
  pb <- NULL;
  for (event.group1.1 in ((CV[[i]]+1):CV[[i+1]])) {
    for (event.group1.2 in 0:(CV[[i+1]] - event.group1.1)) {
      pb <- c(pb, choose(TE[[i]], event.group1.1)*(theta**event.group1.1)*((1-
theta)**(TE[[i]]-event.group1.1)) *
        choose(N[[i+1]], event.group1.2)*(theta**event.group1.2)*((1-
theta)**(N[[i+1]]-event.group1.2)))
    }
  }
  P[[i+1]] <- pb
}
print(c(sum(P[[1]]), sum(P[[2]]), sum(P[[3]]), sum(P[[4]]), sum(P[[5]])))
}
```

#第一種の過誤確率

```
sum(alpha(0.3, 32, 62, 92, 120, 164, 6, 15, 25, 35, 53))
```