

## 非劣性試験 non-inferiority

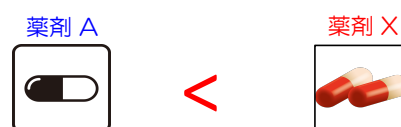
臨床におけるリスク（副作用）とベネフィット（効果）は大きく分けて以下の7つに分類されます。

- ① ～と比較して 効果は大きい、副作用も大きい
- ② ～と比較して 効果は小さい、副作用も小さい
- ③ ～と比較して 効果は小さい、副作用は大きい
- ④ ～と比較して 効果は大きい、副作用は少ない
- ⑤ ～と比較して 効果は同じだが、副作用は少ない
- ⑥ ～と比較して 効果は同じだが、副作用は大きい
- ⑦ ～と比較して 効果も副作用も同じだが 他のメリットがある

この中でリスク・ベネフィットを考えて倫理的に問題があるのは③と⑥でしょう。一方、最も理想的なものは④になります。①は Toxic New として、②、⑤は Less Toxic New として高齢者などの研究に、⑦は安価であるとか、外来通院で可能などが該当し、研究内容により仮説として妥当と考えられます。⑤と⑦の臨床研究は**非劣性試験**として行われています。

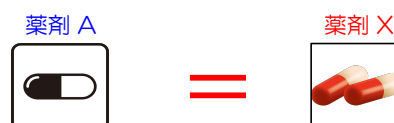
通常臨床研究では「**薬剤 X** の治療効果は **薬剤 A** の治療効果と比較して優れている」という**優越性**を検証しています。

### 優越性



しかしながら、⑤と⑦のような**同等性**の検証が必要な場合もあります。

### 同等性



薬剤Aと薬剤Xの効果は同じ

がん領域では生命延長効果を比較しますので、ハザード比で優越性を検証しますが、ここでは理解のためにまず効果（率）を用いて説明して行きます。最後にハザード比での解説をします。

「薬剤 X の治療効果は 70%、薬剤 A の治療効果を 50%」として比較する **優越性試験**を行う場合には両群で 186 例の症例が必要になります。

必要症例数は比較する差が小さいほど多く必要になります。薬剤 X の治療効果を 60% とすると 776 例必要になり、55% とすると 3,130 例、51% とすると 78,480 例必要になります。**同等性**を検証しようとして 50% にすると天文学的な必要症例数になるでしょう。

### 比較する差が小さいと症例数は多くいる

|                        |      |      |      |      |      |       |
|------------------------|------|------|------|------|------|-------|
| 対照群有効率 ( $\pi_1$ )     | 0.5  | 0.5  | 0.5  | 0.5  | 0.5  | 0.5   |
| 治療群有効率 ( $\pi_2$ )     | 0.7  | 0.65 | 0.6  | 0.55 | 0.53 | 0.51  |
| 症例数の比 ( $n/m = \psi$ ) | 1    | 1    | 1    | 1    | 1    | 1     |
| $\alpha$ エラー*          | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05  |
| Power                  | 0.8  | 0.8  | 0.8  | 0.8  | 0.8  | 0.8   |
| 対照群症例数 (m)**           | 93   | 170  | 388  | 1565 | 4356 | 39240 |
| 治療群症例数 (n)             | 93   | 170  | 388  | 1565 | 4356 | 39240 |
| 総症例数 (N)               | 186  | 340  | 776  | 3130 | 8712 | 78480 |

臨床的に意味があるか？実現可能か？

そこで**非劣性試験**が必要になって来ます。「**有効性は同じだが・・・**」の代わりに「**有効性は劣らないが・・・**」として薬剤 X の治療効果を薬剤 A の治療効果を比較します。

実際には「**有効性は劣らないが・・・**」では無く「**リスクとベネフィットのバランスを考慮して、ベネフィットが大きければ、多少効果が劣っていても良い**」として両者を比較します。

例えば薬剤 A の治療効果は 50%、しかし有害事象が強くと、毎日服用しなくてはならず、1 回の治療費が 50,000 円掛かります。薬剤 X の治療効

### X 治療を A 治療と比べた場合 患者に應用して良いのは次のうちどれか？

- 有効性が高く、安全性も高い 
- 有効性が高いが、安全性は劣る  (toxic new)
- 有効性は劣らないが、安全性は高い  (non-inferiority)
- 有効性は劣らないが、安全性は劣る 
- 有効性は劣るが、安全性は高い  (elderly etc.)
- 有効性が劣るが、安全性も劣る 
- 安全性と有効性も劣らないが、安い  (non-inferiority)

薬剤 A



標準治療薬剤

治療効果：50%

副作用が強い  
毎日服用する  
1錠 50,000円  
・・・

薬剤 X



新しい薬剤

治療効果：45%

ほとんど副作用が無い  
1週間に1度の服用で良い  
1錠 1,000円  
・・・

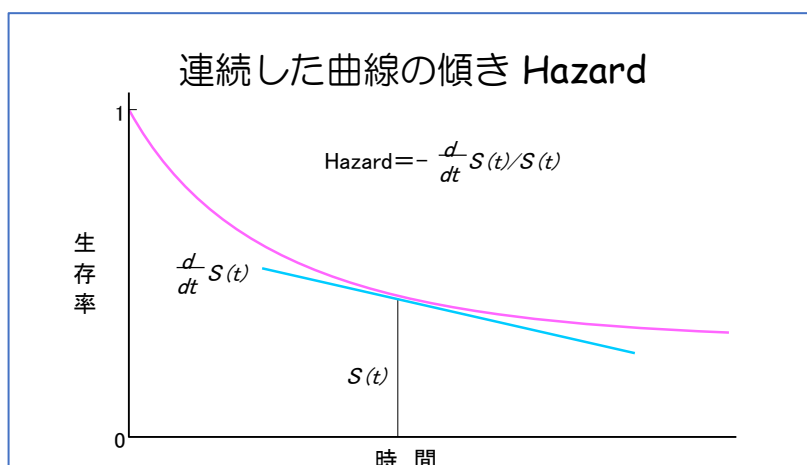
果は 45%と若干低いのですが、ほとんど有害事象が無い、1 週間に 1 度の服用で良く、1 回の治療費は 1,000 円しか掛かりません。5%の治療効果の差程度であれば薬剤 X 治療の選択も良いと考えます。この時の 5%を**非劣性マージン**と呼びます。実際に前向きに両薬剤を比較し、結果的に有効性の差が 5%以内であれば薬剤 X の薬剤 A に対する非劣性が検証されたことになります。

「どの位劣っていても良いのか」は**リスクとベネフィットのバランス**により決めて行きます。もし薬剤 X の効果を 40%と 10%の非劣性マージンを設定するのであれば薬剤 X のベネフィットは更に大きく無くてはなりません。

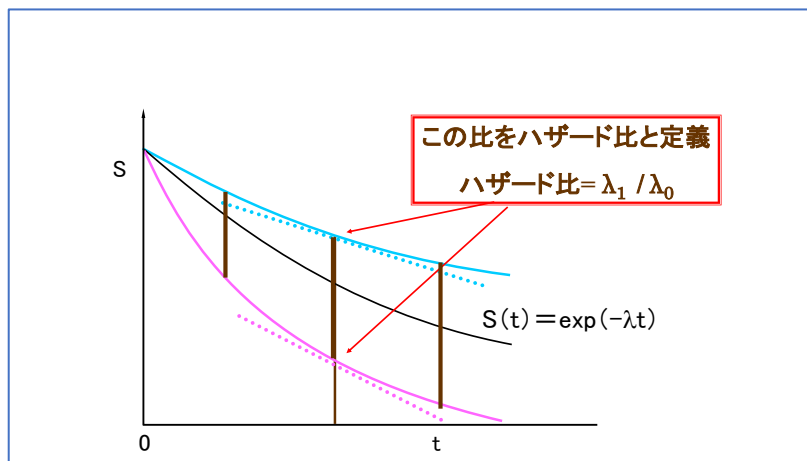
## 生命延長効果の非劣性試験

ここでハザード比の復習です。

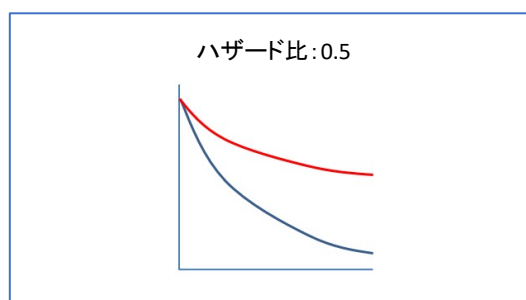
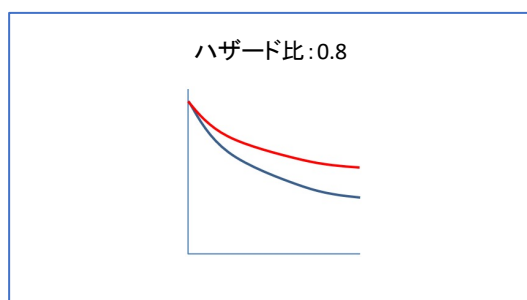
連続した曲線の傾きを Hazard と呼びます。



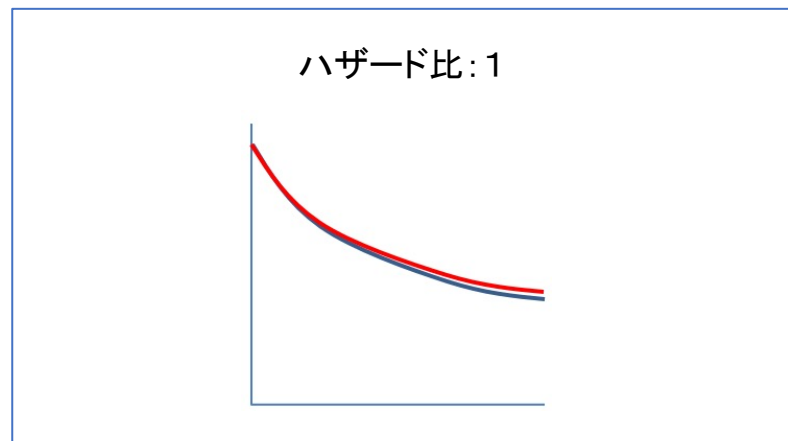
異なる生存曲線の傾きの「比」をハザード比と定義します。



生存曲線の差が大きいほどハザード比は小さくなります。



生存曲線の差が無い場合にはハザード比は「1」になります。(=同等)



非劣性試験ではハザード比が「1」に近いほど「劣らない」と判断されるのですが、それではどの程度までならリスクとベネフィットのバランスを考えて許容されるかによって「非劣性マージン」を決定します。(ここが最も理解しにくいところです)

一般的には 1.25 あるいは 1.3 と設定している試験が多いようです。

### 非劣性試験の症例数算定

SWOG の Statistical Tools の Survival Noninferiority を利用しましょう。





STATISTICAL TOOLS | DESIGN | ANALYSIS | PROBABILITIES | OTHER TOOLS | ABOUT US

### Survival Noninferiority

Survival Noninferiority is a program to calculate either estimates of sample size or power for equivalence studies. Exponential distributions are assumed and are specified using the ratio of the hazard for the experimental treatment to the hazard for the standard treatment. The ratio of hazard rates describing equivalence is defined in the same way.

| User Input                                                                  |              | Program Output                     |
|-----------------------------------------------------------------------------|--------------|------------------------------------|
| <input checked="" type="radio"/> Sample size<br><input type="radio"/> Power | 1 Sided Test | Alpha<br>.05                       |
| Accrual                                                                     | Follow       | Comp Risk<br>0.00                  |
| Proportion in Experimental Sample<br>.5                                     |              |                                    |
| Hazard Rate (Experimental)                                                  | Hazard Ratio | Ratio Defining Equivalence<br>1.20 |
| Power<br>.90                                                                | Sample Size  |                                    |

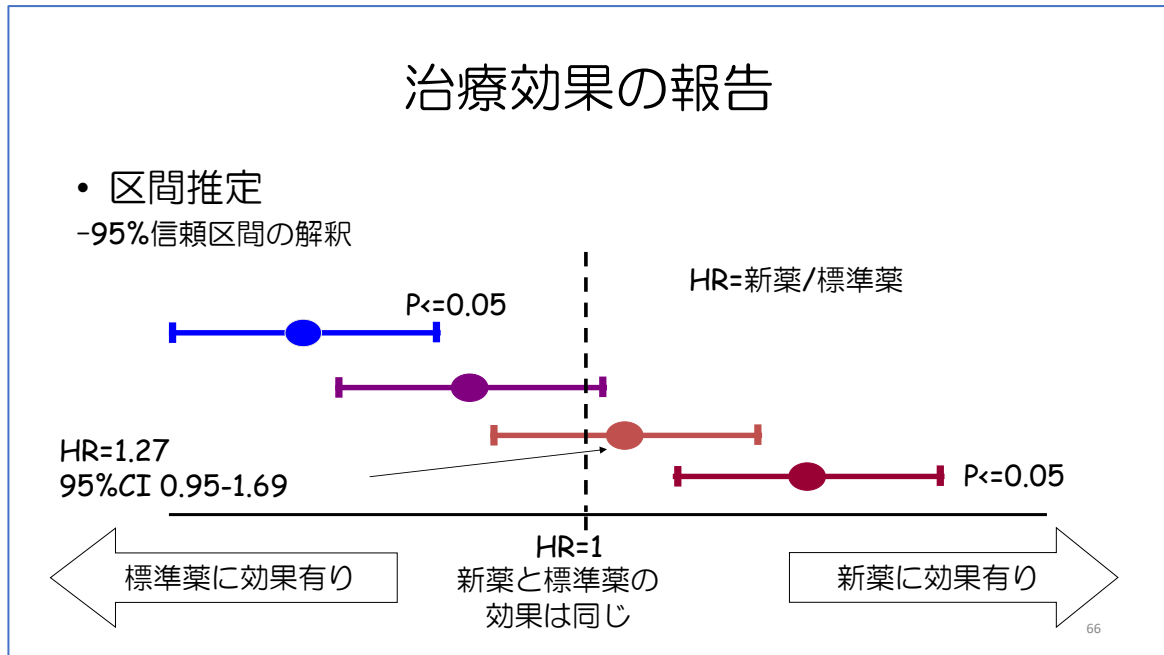
Calculate

[Help Document](#)

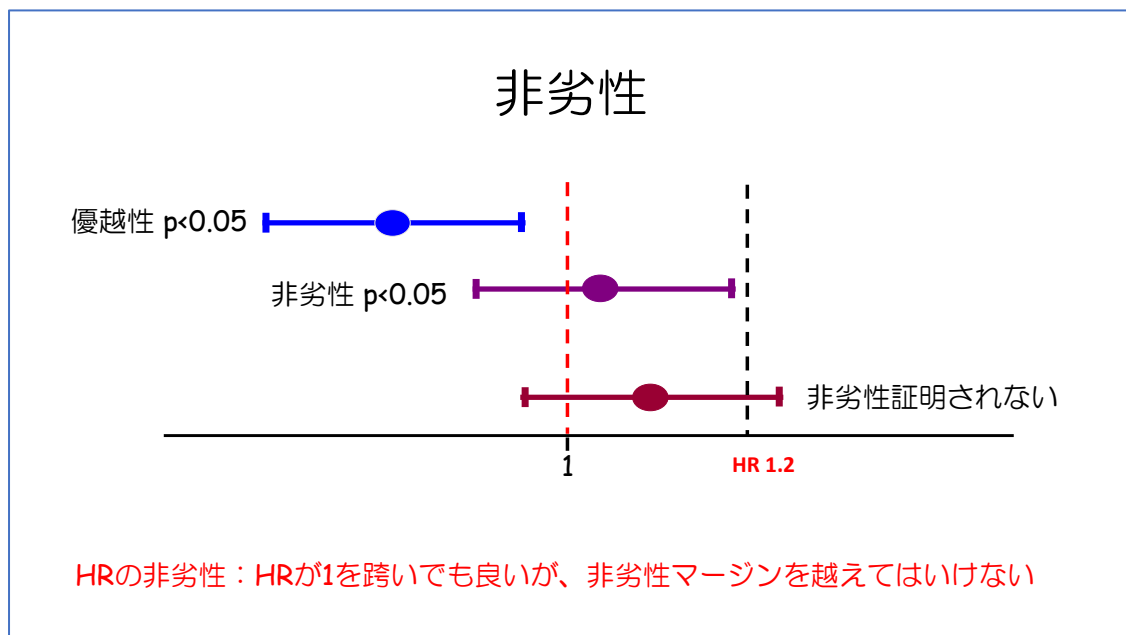
## ハザード比による検定

優越性を検定する場合にはハザード比 (Hazard Ratio: HR) の 95%信頼区間が「1」を跨いではいけません。標準薬と新薬で行った第3相試験のハザード比が  $HR=1.27$  (95% CI: 0.95-1.69) だった場合は両群に差がありません。

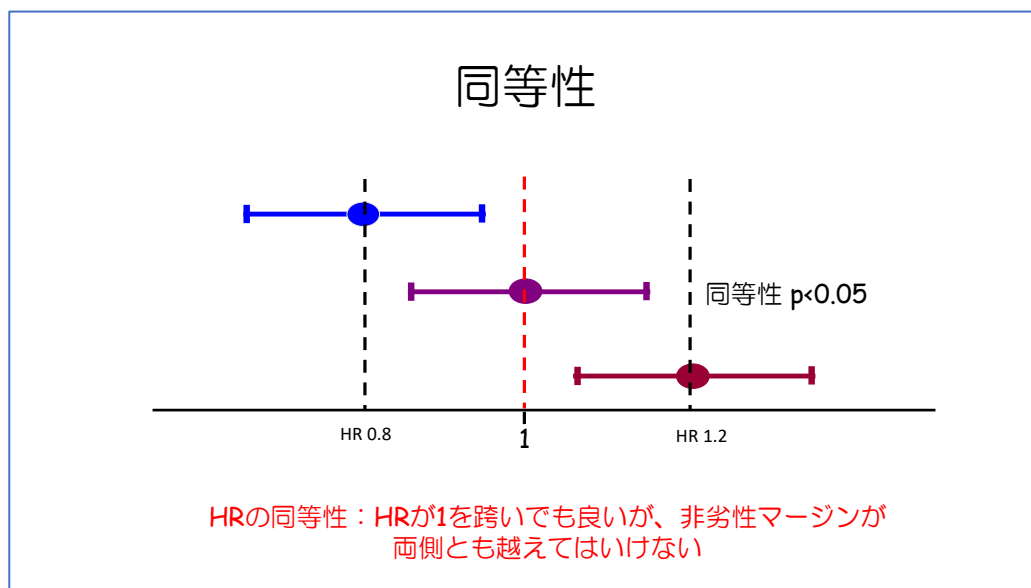
HR の 95%信頼区間が「1」を跨がなければ新薬に効果ありと判断されます。



非劣性試験の場合、HR の 95%信頼区間が非劣性マージンを跨がなければ非劣性を証明できます。



同等性を検証しようとする場合（実際にはほとんど行われませんが）、同等性試験では両側に非劣性マージンを設定し。試験結果の HR の 95%信頼区間の上限、下限とも非劣性マージンを跨がなければ同等性が検証されたことになります。



#### Non-inferiority 試験の例：SOFT 試験

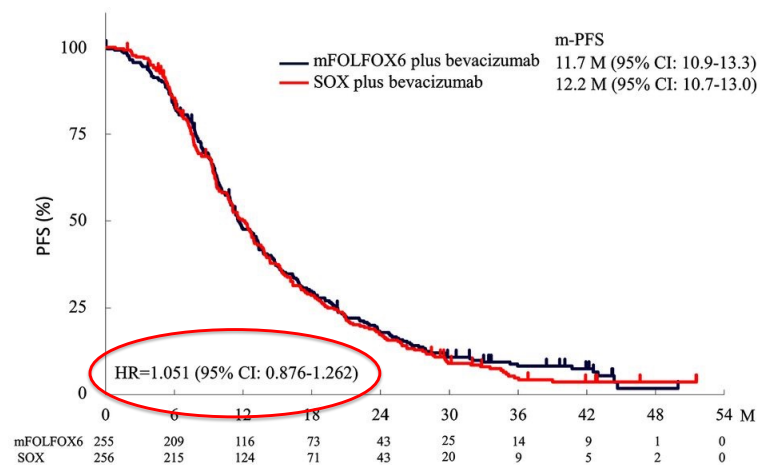
**ESMO** *Open*  
*Cancer Horizons*

search Q

Advanced Search

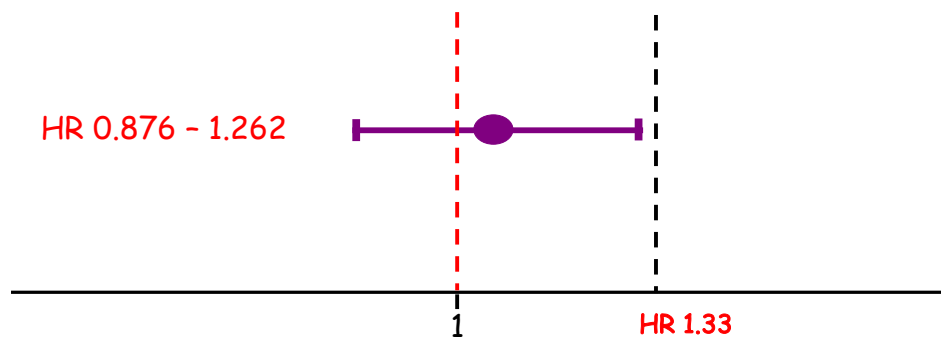
③ S-1 and oxaliplatin (SOX) plus bevacizumab versus mFOLFOX6 plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer: updated overall survival analyses of the open-label, non-inferiority, randomised phase III: SOFT study

- ・ 進行再発大腸癌の一次治療は FOLFOX である。
- ・ S-1 は経口投与可能であり利便性が高い
- ・ S-1/Oxaliplatin (SOX) も一次治療として有用性が示唆された
- ・ 本研究では SOX の FOLFOX に対する無増悪期間の非劣性試験を行った
- ・ 非劣性マージンは HR を 1.33 未満とした
- ・ 必要症例数は各群 250 例ずつとした



## 非劣性の証明

非劣性  $p < 0.05$



結果として SOX/Bmab の mFOLFOX/Bmab に対する非劣性は検証された。

この試験で設定された非劣性マージン 1.33 に対して批判もあったが（一般的には 1.2 程度）、両群の生存曲線の重なり、PFS を主要評価項目とした場合に再発確認は 1 ヶ月間隔であること、非劣性マージンを 1.2 に設定した場合の試験実現可能性から容認されている。

## Non-inferiority 試験の例：IDEA 試験

# The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

MARCH 29, 2018

VOL. 378 NO. 13

## Duration of Adjuvant Chemotherapy for Stage III Colon Cancer

A. Grothey, A.F. Sobrero, A.F. Shields, T. Yoshino, J. Paul, J. Taieb, J. Souglakos, Q. Shi, R. Kerr, R. Labianca, J.A. Meyerhardt, D. Vernerey, T. Yamanaka, I. Boukovinas, J.P. Meyers, L.A. Renfro, D. Niedzwiecki, T. Watanabe,\* V. Torri, M. Saunders, D.J. Sargent,\* T. Andre, and T. Iveson

### 研究の背景

大腸癌 Stage III の術後補助化学療法の標準治療は FOLFOX あるいは CAPOX の 6 ヶ月投与である。しかしながら、oxaliplatin には特有の神経毒性があり短期間投与に変更可能か検討する。

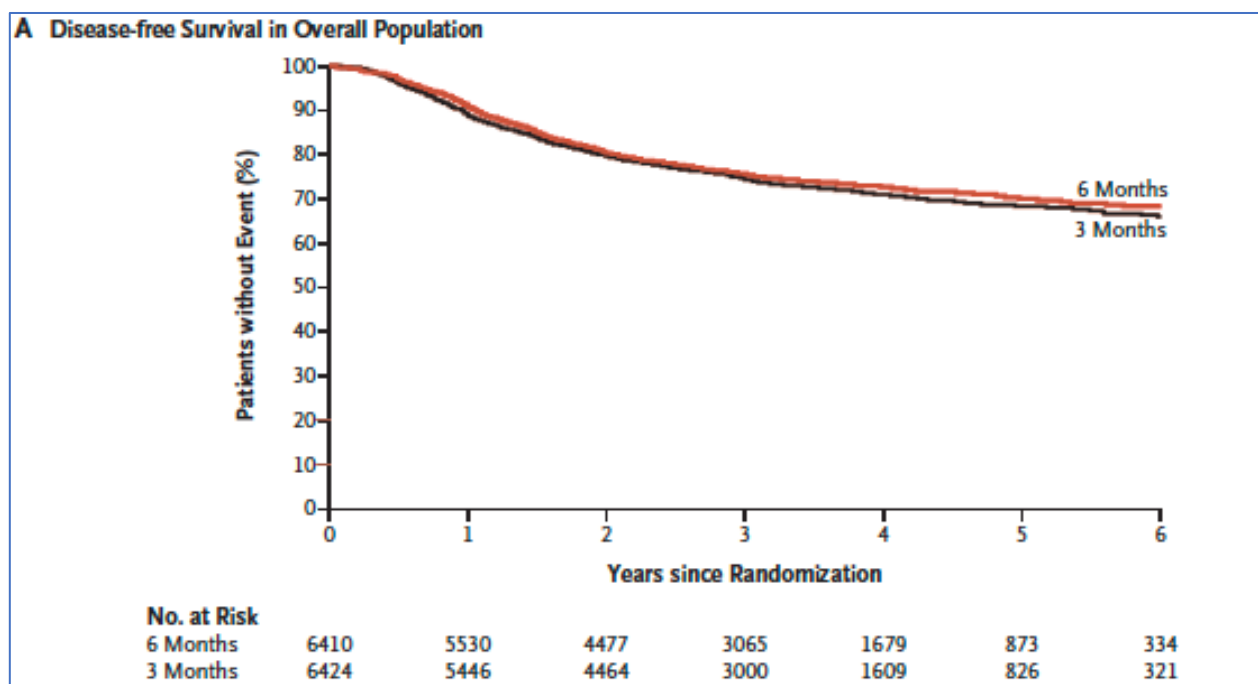
### 方法

6 ヶ月投与と 3 ヶ月投与を比較する。主要評価項目は 3 年無再発率。非劣性マージンは 1.12 とした。

### 結果

12,384 症例を集積した。3 年無再発におけるハザード比は 1.07、95%CI 1.00-1.15)。HR の 95%CI が非劣性マージンの 1.12 を跨いだので結果は negative。

下図のようにほぼ生存曲線が重なっているように見えるが統計学的には Negative。





## Risk-Benefit 評価 さらに！

- 有効性が高く、安全性も高い →しかし超高額
- 有効性が高いが、安全性は非常に劣る  
→しかし非常に有効性が高い
- 有効性は劣るが、安全性は高い  
→どの位有効性が劣るか？

## Inferiority Non-inferiority

- 安全性と有効性も劣らないが、安い
- 安全性と有効性も劣らないが、便利（注射：経口）
- 安全性と有効性も劣らないが、便利（毎日：週一）
- 安全性と有効性も劣らないが、便利（入院：外来）
- 安全性と有効性も劣らないが、痛み（痛い：痛くない）
- 安全性と有効性も劣らないが、.....