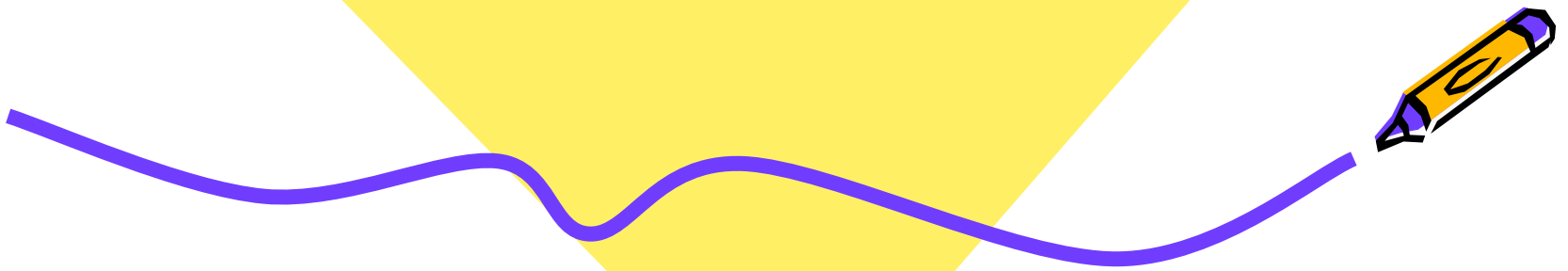




第II相臨床試験  
癌患者さんの試験を中心に



## 薬剤に関する第Ⅱ相臨床試験

- 初期用量は第Ⅰ相臨床試験の結果より設定する。
- 用量は通常増量しない(デザインによる);むしろ1人でも患者さんに毒性を認めたら誰も毒性を示さないところまで減量する。

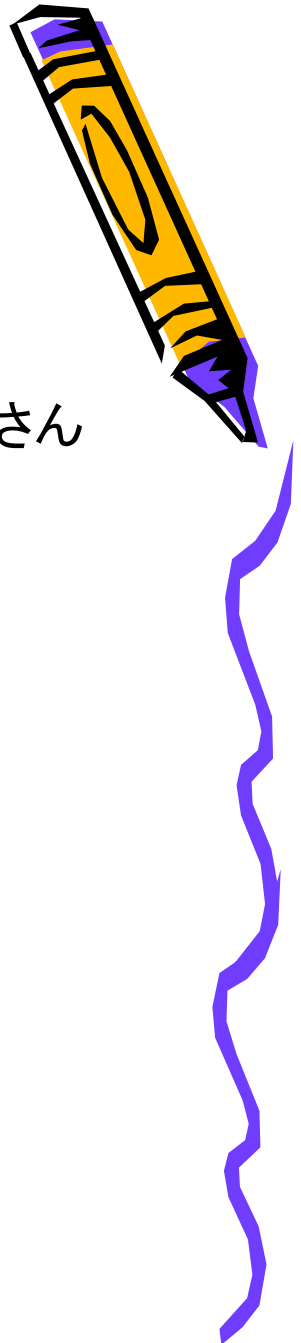




- ある程度反応の得られそうな対象を選ぶ。そうすることで、その治療の可能性を摘んでしまわない。
- 薬剤に対する反応性、その他の効果が第Ⅱ相臨床試験における主な目的であるが、毒性や副作用についての情報収集も重要である。
- 第Ⅰ相臨床試験とはすこし異なった患者層を対象とする。健丈人を対象とせず、薬剤の対象疾患を持つ患者さんを対象とする。しかしながら希な病気に対しては健丈人を対象に行なわれることもある。薬剤の反応性を測定可能な病気を対象とする。患者さんの余命が薬剤の反応をみるのに十分な時間あることを必要とする。
- 通常はその薬剤試験の対象疾患は1つに絞られる。
- 経費と時間が限られる。しかも他社との競合等も存在し、複雑な世界である。また第Ⅲ相臨床試験は高額のコストがかかるため、第Ⅱ相臨床試験で十分な自信を得て次のステップにすすむ必要がある。しかし疾患および薬剤の性格によっては、観察期間が短かったり、用量設定がうまくいかなかったり、健康すぎる、あるいは病気が重過ぎて本当の効果をうまく測定できないことも十分ありえる。



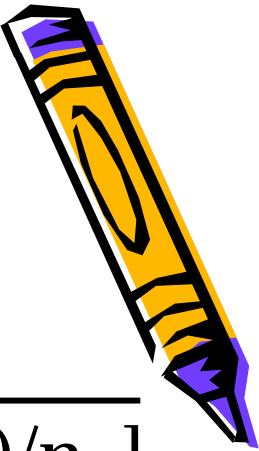
- そのようなわけで、倫理的に許されるならば、状態のよい患者さんを対象に、前治療の少ないものを選ぶ。
- また、結果を測定できることが前提である。





- 他に良い治療の存在しない病気のステージの患者さんの場合（不治の病）、まず14—20人を対象に薬を投与する。例：14人に投与し全く効果を認めなければその時点で中止する。なぜなら真の反応率が20%だとした場合、14人中誰も反応しない確率は0.05以下であり統計学的に起こりにくいことである（14人に反応がみられなければ、本当は20%も反応率が期待できる治療であることは統計学的には考えにくい）。もしも、少なくとも1人が薬剤に反応したら、新たに10—25人の患者さんに薬剤を投与し、反応率と信頼区間を算出する。よって20-80人位の患者さんを対象とする。通常は100-200人を超えることはない。それでは95%CIはどのように算出するのか（図1）。




$$(\pi_B - \pi_A) \pm 1.96\sqrt{[\pi_A(1 - \pi_A)/n_A + \pi_B(1 - \pi_B)/n_B]}$$

$\pi_A$  = probability of response rate by treatment A

$\pi_B$  = probability of response rate by treatment B

$n_A$  = number of sample treated by A

$n_B$  = number of sample treated by B

図1 一般的な95%信頼区間の算出公式



- 95%の確率で、観察した反応率が真の反応率10%前後にしたい (図2)。

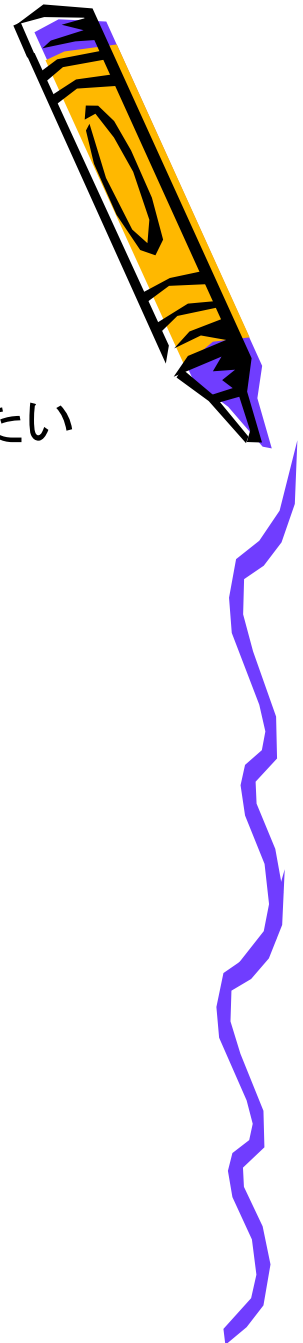
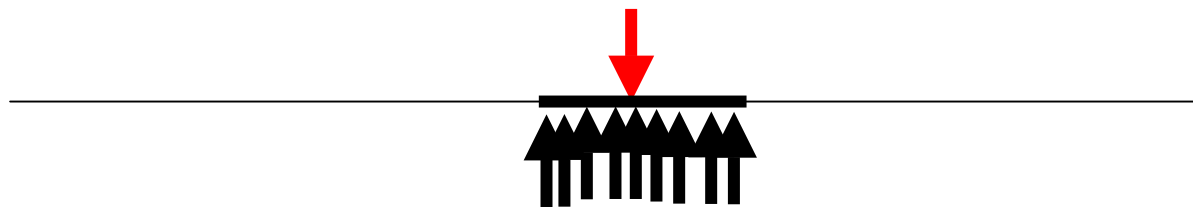


図2: phase II trial (1 arm) において、95%CIを確認したいと思う。95%の確率で、観察した反応率が真の反応率の10%前後になるとすれば、何人の患者さんでその治療を試す必要があるか？反応率を仮想して、それぞれ値を算出する。

真の反応率



研究で得られた反応率が真の反応率の前後10%になるようにしたい。100回同じことをくり返したら95回は真の反応率の前後10%に収まるということ。





- 何人の患者さんでその治療を試す必要があるか(図3)? 20%の奏効率に10%の範囲で近づけるためには62人の対象が必要となる。他にも治療薬が存在する場合、短期間に第Ⅱ相臨床試験を終える。



反応率を $\pi$ と設定すると。。。

one arm なので、

$$1.96 \sqrt{[\pi (1 - \pi)/n]} = 0.1$$

When  $\pi = 0.8$ ,  $n = 62$

When  $\pi = 0.7$ ,  $n = 81$

When  $\pi = 0.6$ ,  $n = 93$

When  $\pi = 0.5$ ,  $n = 96$

When  $\pi = 0.4$ ,  $n = 93$

When  $\pi = 0.3$ ,  $n = 81$

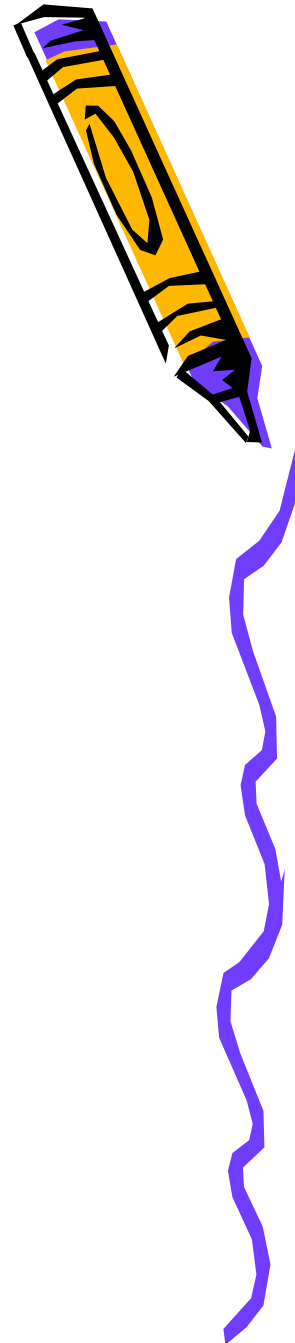
When  $\pi = 0.2$ ,  $n = 62$

When  $\pi = 0.1$ ,  $n = 35$

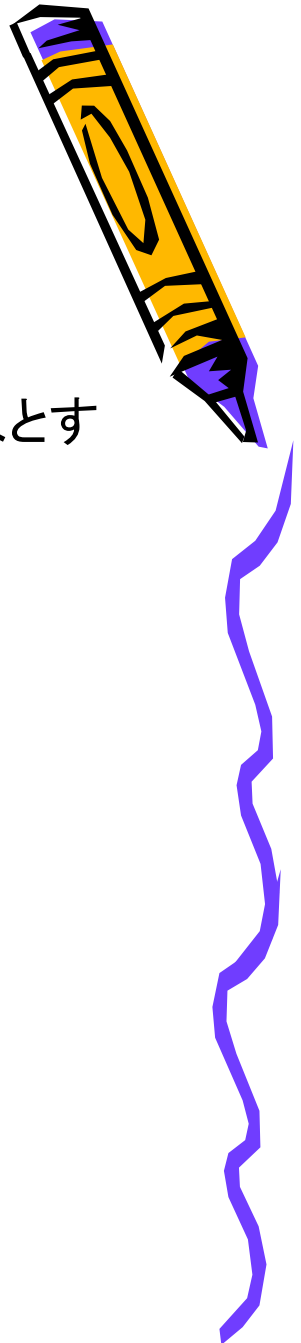
となる。反応率が 50%のときに最も多い人数を必要とする。



図3: 解答



- よって、反応率55%をもって有効と考える。42人の患者を対象とすれば、95%信頼区間を55%前後15%にすることができる。



- ....Therefore, achievement of a response rate of more than 55% was considered treatment results worthy of further development of this dose and schedule of rituximab. Accrual of 42 assessable patients to this trial allowed adequate assessment of treatment toxicity and provided a response rate with 95% confidence intervals of plus minus 15%.



Modified: J Clin Oncol 2003; 21: 1746-51.

- $P=0.55$ であったときの95%信頼区間は上記公式で得られる。その結果、42人対象として必要であるという計算が成り立つ。



$$P \pm 1.96 \sqrt{\frac{P(1-P)}{N}}$$

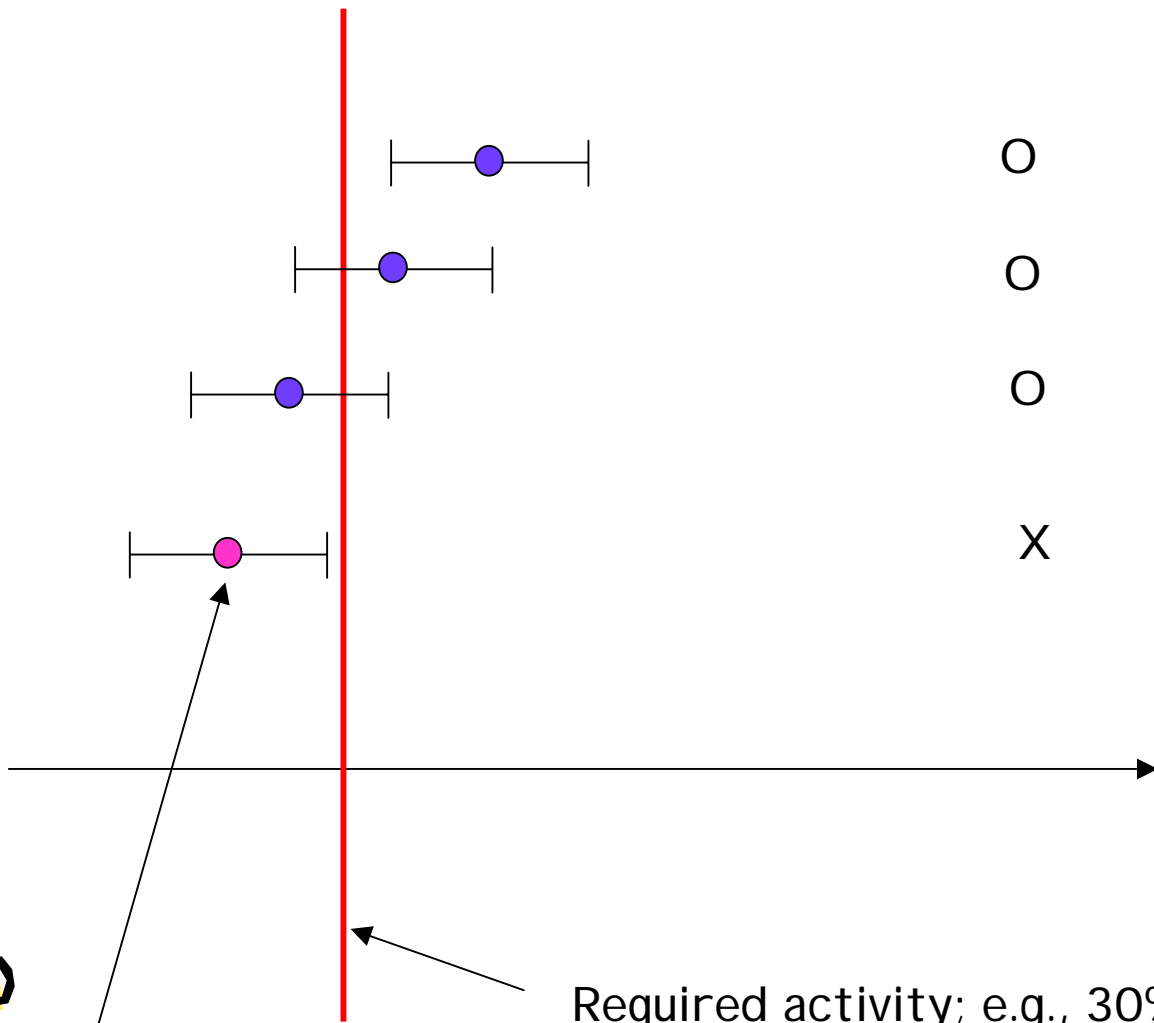
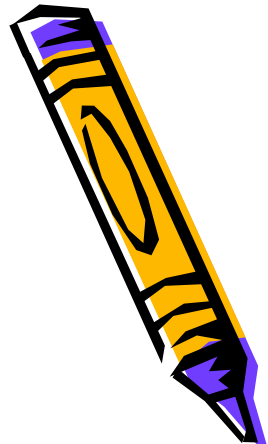
$$0.15 = 1.96 \sqrt{\frac{0.55(1-0.55)}{N}}$$

N = 42; but we should recruit a little bit more patients to consider drop out of intensified therapy.



- 疾患Aに対して、標準治療では約30%の反応率が得られているとしよう。もしも、新治療Bを用いて治療したときある反応率が得られる。しかし、反応率が95%の確率で30%に達しないとすれば、標準治療より劣ると考えられるため、臨床試験は中止する。





Required activity; e.g., 30%

The treatment would be rejected as of little interest.





- 新治療Cに相当する標準治療Dの反応率は20%だと仮定する。統計ソフトSTATA (<http://www.stata.com>) を用いて、10人に新治療Cを適応して、1人も反応がなかったとする。真の反応率(たまたま最初の10人で反応は得られなかったが、10000人やれば2000人くらいが反応する)が20%だが、たまたま最初の10人に反応がみられなかったということで棄却されてしまう確率は、10%である。もう少し人数を増やして検証するか、その時点であきらめるかは、研究者次第である。



Binomial distribution  
Immediate test

Stata Command

bitesti 10 0 .2

Patients enrolled

Success

True probability

```
. bitesti 10 0 .2
      N   Observed k   Expected k   Assumed p   Observed p
-----
      10         0         2         0.20000   0.00000
Pr(k >= 0)           = 1.000000 (one-sided test)
Pr(k <= 0)           = 0.107374 (one-sided test)
Pr(k <= 0 or k >= 4) = 0.228248 (two-sided test)
```

Rejection error ( $\beta$ ) (%) = 10.7%



- 簡単な早見表である。例えば、標準治療が30%の反応率を得ているとする。もしも9人に新しい治療を行って、誰も反応しなかった場合、その新しい治療が標準薬の治療成績を越えるということは95%の確率で否定できる。



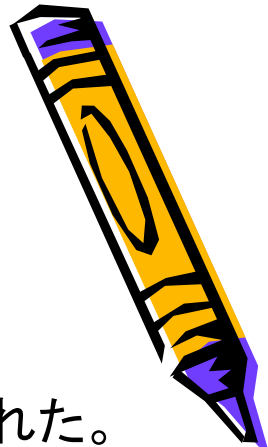


	Therapeutic Effectiveness (%)									
$\beta$ (%)	5	10	15	20	25	30	35	40	45	50
5	59	29	19	14	11	9	7	6	6	5
10	45	22	15	11	9	7	6	5	4	4

For example, if the study produced no responses among the first 9 patients, one would have 95% confidence in rejecting the treatment as being less than 30% active.



- 95%信頼区間でしてみる。例えば、20人中6人で反応がみられた。しかし人数が少ないので、次回3人しか反応がないかもしれないし、10人が反応するかもしれない。統計計算すると、今回の反応率は30%であるが、95%信頼区間は12%から54%と幅広い。しかし、今後この研究を何度か行ったとしても、10%をきる可能性は低いと言える。



95% CI

```
Stata Command  
cii 20 6
```

In 20 patients, 6 will succeed.



```
. cii 20 6
```

Variable	Obs	Mean	Std. Err.	— Binomial Exact — [95% Conf. Interval]	
	20	.3	.1024695	.1189316	.5427892

With 95% confidence, true response rate is more than 10%.



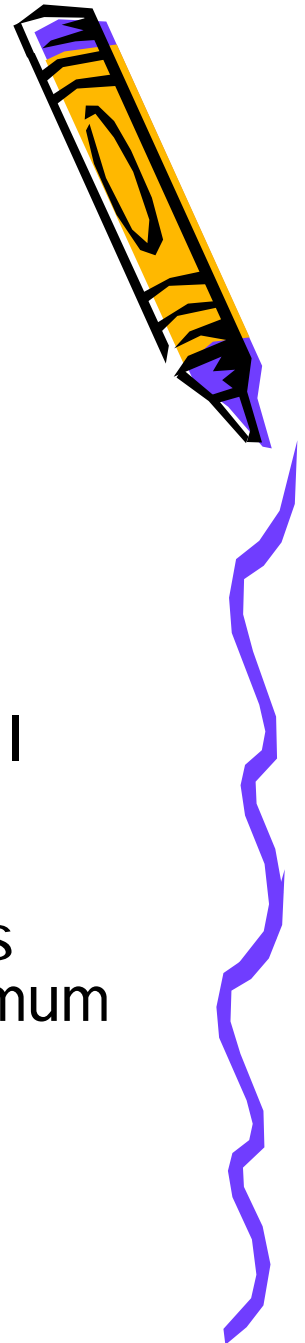


- それでは、最初の9人で1人が反応を示したとしよう。その次はどうすべきか？
- まだ、95%信頼区間が広いので、もう少しせまくなるように対象人数を増やす。
- そうすれば、第III相臨床試験に移行するかどうかの判断もつきやすい。
- 一般的には35人以上でphase II clinical trial を行う。



If more than one response is seen among the first 9 patients...

- The accrual is continued until a study group is reached that is large enough to provide an estimate of response rate with a specified precision.
- This estimate will be helpful in planning the further investment of resources in a phase III study and in determining whether this step is justified.
- The minimum total size for a phase II study is typically 35 subjects, which allows for a maximum standard error of approximately 8%.







- 例えばがん患者に対する新しい抗がん剤治療を考える。そして、腫瘍が縮小するか否かで判定する。対象人数をN, 腫瘍の縮小をみるものをS, その割合を $S/N$ とする。
- そして、2つのハードルを定める。1つは標準治療の腫瘍縮小率で、そこを超えることができなければ新しい治療の効果を見出せないとしてあきらめる。もう1つは、新しい治療として期待されている腫瘍縮小率で、高いレベルを設定する。

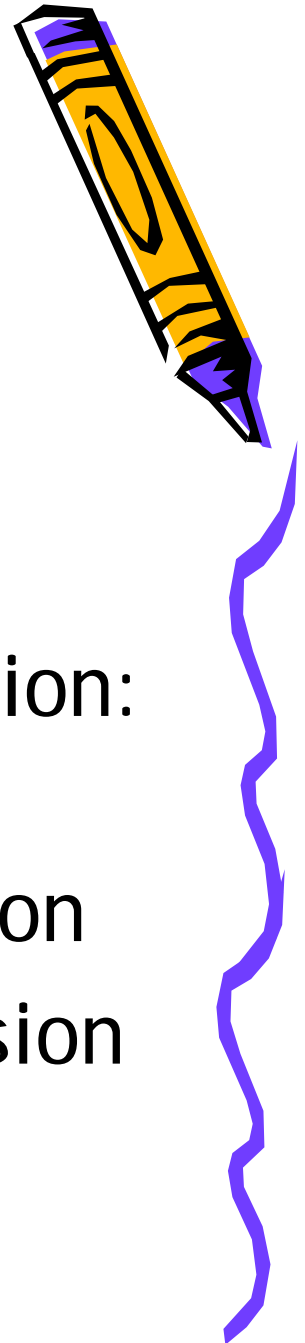


# Evaluation of tumor regression

- N: evaluable patients accrue
- S: tumor regression
- P: estimated probability of regression: S/N
- $P_0$ : required probability of regression
- $P_A$ : expected probability of regression

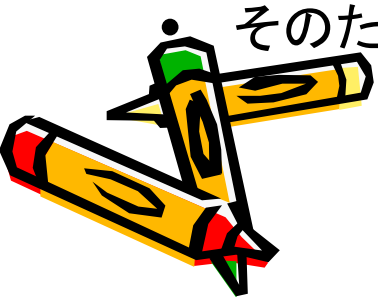
$$H_0: P \leq P_0$$

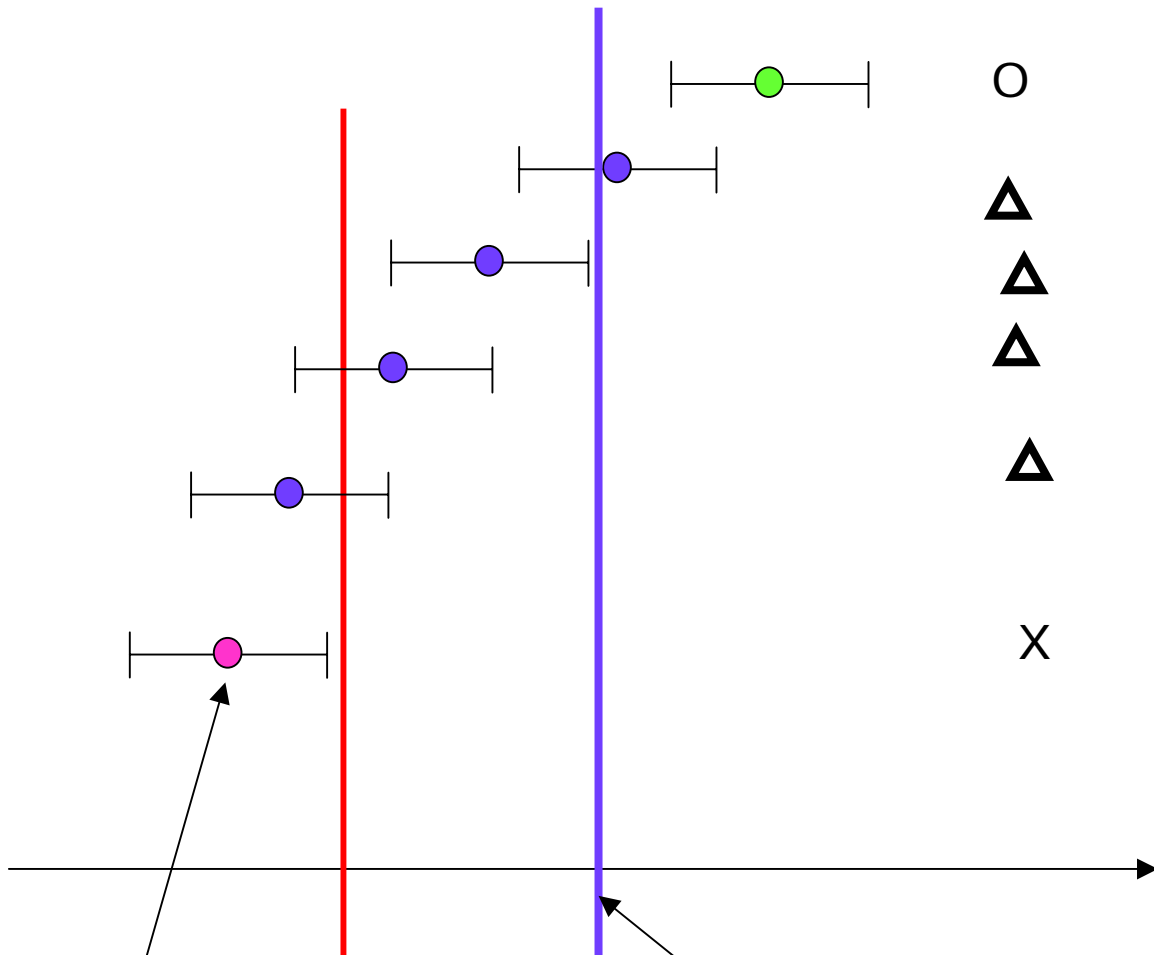
$$H_A: P_A \leq P$$





- 例えば、第一のハードルを10%に設定する。そして、ここで気にするのは、本当は10%も治療効果がないのにあるとい結論してしまうことである。すなわち、アルファエラーである。
- 例えば第一のハードルを越えて、少なくとも10よりは効くということが判ったとする。しかし、有効率が10%から100%というのでは、次の計画を立てようがない。そこで、10%有効なのか、30%有効なのか識別できることを目標にしよう。このことは、真の有効率が20%であっても10%と識別できないということである。つまり、10%だろうが20%だろうが気にしない。しかし、30%効いている治療であるならば、みすみす第二相試験で中止したくない。ということである。当然、30%より20%の方が必要対象数が大きくなる。
- そのためには何人必要か？

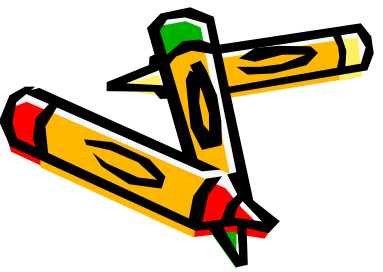




Required activity; e.g., 10%;  $\alpha = 5\%$

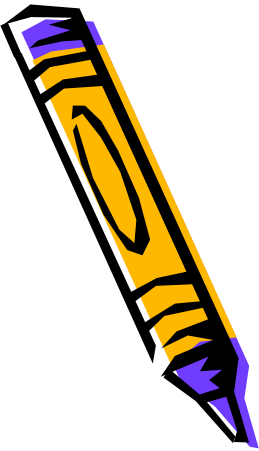
Expected activity; e.g., 30%;  $\beta = 20\%$

The treatment would be rejected as of little interest.



- 先の例を具体的に計算してみる。アルファエラー:5%、パワー:80%に設定すると、必然的にZ値はそれぞれ1.96, 0.84となる。公式に当てはめると、対象人数として24人が妥当という結果となる。




$$N = \frac{\{[Z_{\alpha} \times P_0(1-P_0)]^{1/2} + [Z_{\beta} \times P_A(1-P_A)]^{1/2}\}^2}{[P_0 - P_A]^2}$$

Assuming  $Z_{\alpha} = 1.96$  ( $\alpha = 0.05$ ),  $Z_{\beta} = 0.84$  ( $\beta = 0.2$ )

$$N = \frac{\{1.96 \times [0.1(1-0.1)]^{1/2} + 0.84 \times [0.3(1-0.3)]^{1/2}\}^2}{[0.1 - 0.3]^2}$$

= 23.5 < 24 participants will be required.



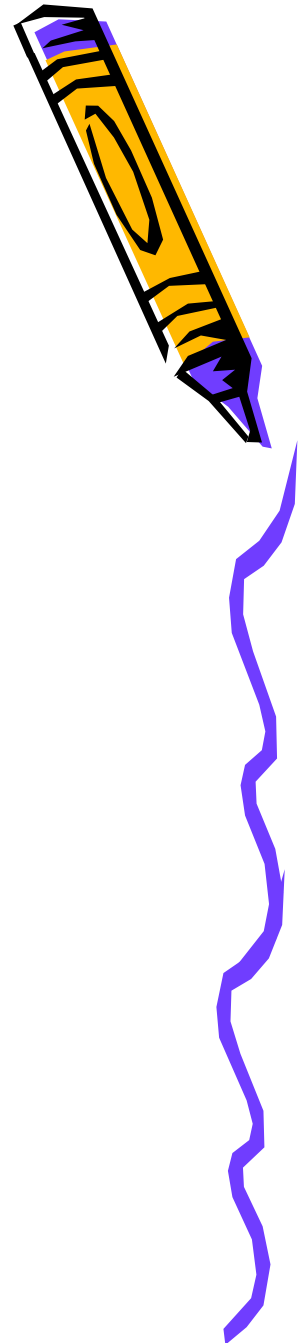
$$S \geq \underbrace{\{N \times P_0 + Z_\alpha [NP_0(1-P_0)]^{1/2}\}}_{\text{Nearest integer of } \{x\}} + 1$$

$$\begin{aligned} & 24 \times 0.1 + 1.96[24 \times 0.1(0.9)]^{1/2} + 1 \\ & = 2.4 + 2.88 + 1 = 6 \end{aligned}$$

= 6 responses are required to pass  
in 24 patients.



- さらに、24人中6人で腫瘍の縮小を認める必要がある。

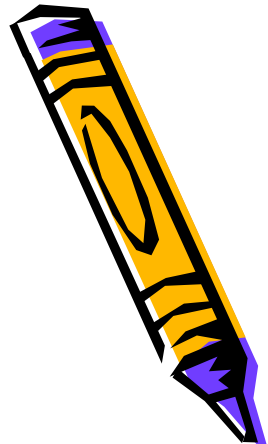






- 再度STATAを用いて検証する。新しい治療の真の腫瘍縮小率が10%であるとき、24人に新しい治療を施せば理論上2.4人が反応する。24人中6人以上反応するということは、真の反応率が10%と考えたときありえないくらいよい反応率ということになる( $p=0.027$ )。一方、新しい治療の真の腫瘍縮小率が30%であるとき、24人に新しい治療を施せば理論上7.2人が反応する。24人中6人以上反応するということは、真の反応率が30%と考えたとき十分ありえる数値である( $p=0.77$ )。





```
. bitesti 24 6 .1
```

N	Observed k	Expected k	Assumed p	Observed p
24	6	2.4	0.10000	0.25000

Pr(k >= 6) = 0.027658 (one-sided test)

$\alpha$  error = 0.028

```
. bitesti 24 6 .3
```

N	Observed k	Expected k	Assumed p	Observed p
24	6	7.2	0.30000	0.25000

Pr(k >= 6) = 0.771192 (one-sided test)

Power (1 -  $\beta$  error) = 0.771



- 早見表である。例えば、20%はどうしてもクリアしたい、できれば40%  
いっているかみたいとき、25人に新しい治療を施行し、9人が反応  
すれば、95%の確率で20%は超えており、73%の確率で40%を超  
えるであろう。



```
. bitesti 25 9 .2
      N   Observed k   Expected k   Assumed p   Observed p
-----
      25           9           5         0.20000     0.36000
Pr(k >= 9)           = 0.046774 (one-sided test)
```

$P_0$	$P_A$	N	$R^* + 1$	a	1-b
0.20	0.40	25	9	0.0468	0.7265
0.20	0.40	30	11	0.0256	0.7068
0.20	0.40	35	12	0.0343	0.8049
0.20	0.40	40	13	0.0431	0.8715
0.20	0.40	45	14	0.0521	0.9163
0.20	0.40	50	16	0.0308	0.9049
0.25	0.50	20	9	0.0410	0.7483
0.25	0.50	25	11	0.0297	0.7878
0.25	0.50	30	12	0.0506	0.8997
0.25	0.50	35	14	0.0363	0.9213

```
. bitesti 35 14 .5
      N   Observed k   Expected k   Assumed p   Observed p
-----
      35          14          17.5        0.50000     0.40000
Pr(k >= 14)           = 0.912267 (one-sided test)
```



- 2ステージ法による対象数の計算。この治療法が5%を切るようなら即刻あきらめ、20%を超えるようなら2nd ステージに進む。具体的には、最初の14人に本治療を施行し、誰も反応しなければ中止する。少なくとも1人が反応したら更に16人に試みる(2<sup>nd</sup> ステージ)。30人に対して20%の反応率をみるとすると、パワー86%、アルファエラー5%となる。





# A two-stage sample size design

- The response rate of interest was 20% or greater, and a response rate of less than 5% would be interpreted as poor efficacy. A lack of any response in the first 14 response-assessable patients would have resulted in termination of the study. Because at least one response occurred in the first 14 response-assessable patients, an additional 16 response-assessable patients were recruited, for a planned total of 30 patients. The estimated power of this design is 0.86, when the true response probability is 20%, given a type I error of 0.057.



J Clin Oncol 2003; 21: 1740-5.



- まず1st ステージ。14人に治療して誰も反応しなかった場合、真の確率が20%である確率は5%未満である。そのため、最初の14人に試して、誰も反応を示さなければ、この段階で臨床試験は中止となる。



A lack of any response in the first 14 response-assessable patients would have resulted in termination of the study.



```
. bitesti 14 0 .2
```

N	Observed k	Expected k	Assumed p	Observed p
14	0	2.8	0.20000	0.00000

Pr(k >= 0)	=	1.000000	(one-sided test)
Pr(k <= 0)	=	0.043980	(one-sided test)
Pr(k <= 0 or k >= 6)	=	0.087835	(two-sided test)

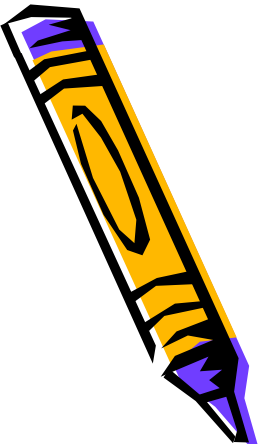






- 10人中3人で有効な治療の真の反応率は7%から65%である。即ち、真の反応率は95%の確率で7%から65%の間に落ち着くことが予想される。あるいは、次に新たな10人に対して治療を行い反応率をみると、7%から65%の間になる。つまり、最初の10人では3人に効いたわけであるが、次の10人での反応率は1人かもしれないし、7人かもしれないということである。
- 対象数を増やせば増やすほど、95%信頼区間の幅が狭くなる。それだけ、真の反応率に近付いている、あるいは次回新たな同じ人数で行うと近い反応率を示すだろう。10万人に対して行えば、ほぼ真の反応率が得られている。




$$95\% \text{ CI} = \pm \sqrt{p(1-p)/N}$$


10人中3人で有効: 95%CI 0.07 – 0.65

100人中30人で有効: 95%CI 0.21 – 0.40

1000人中300人で有効: 95%CI 0.27 – 0.33

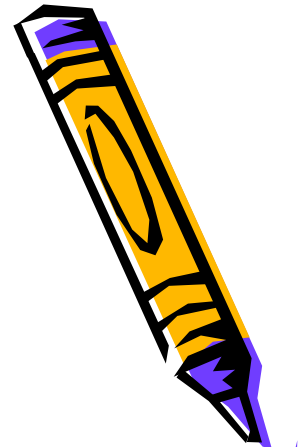
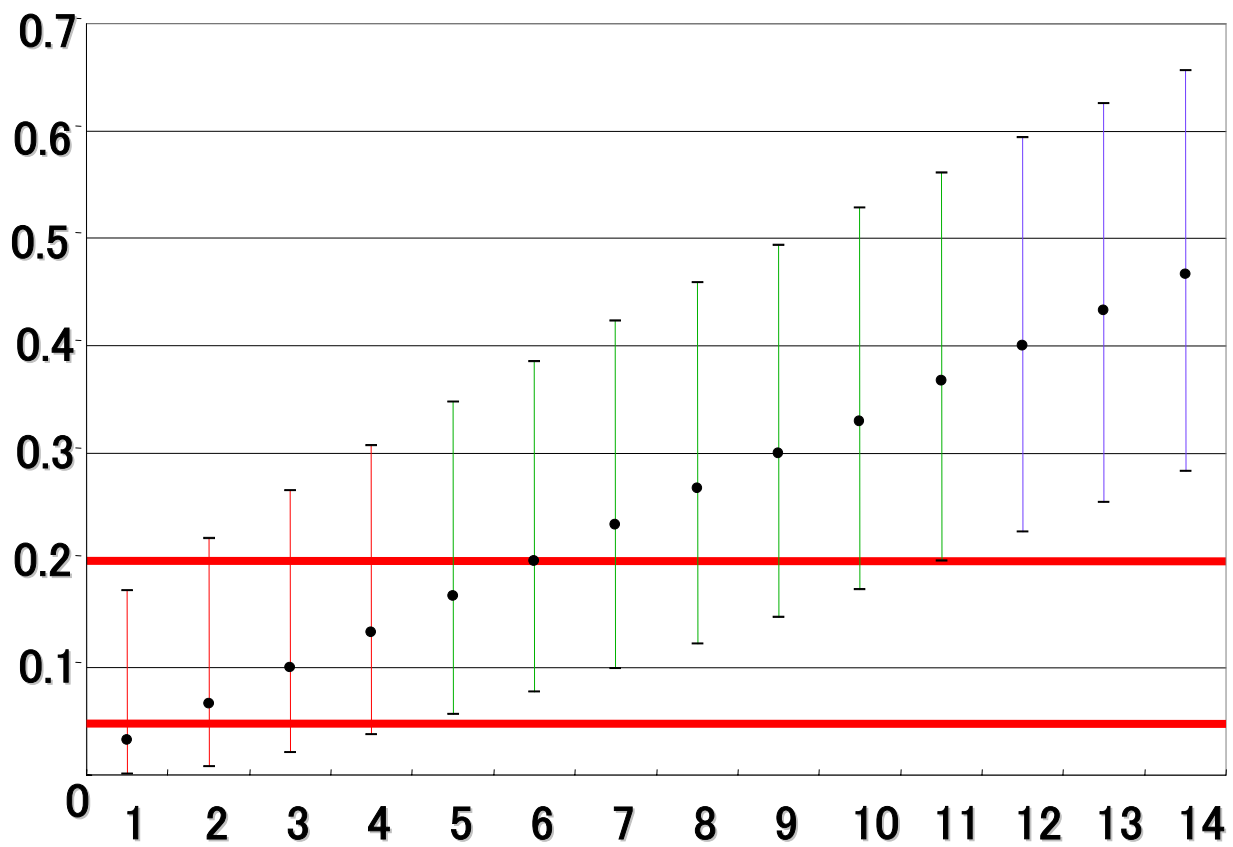
10000人中3000人で有効: 95%CI 0.29 – 0.31

100000人中30000人で有効: 95%CI 0.30 – 0.30



- 14人中1人以上が反応したときの95%信頼区間を示してある。14人中12人に反応が認められると、本治療の反応率は95%の確信をもって20%以上であると結論できる。





- STATAを使用した対象数計算である。5%は少なくとも超え、できれば20%を超える治療法であるかどうかを判定する。対象数を30人に設定すると、アルファエラー5%、パワー84%となる。



```
. sampsi 0.05 0.2, n(30) onesample
```

Estimated power for one-sample comparison of proportion  
to hypothesized value

Test Ho:  $p = 0.0500$ , where  $p$  is the proportion in the population

Assumptions:

```
      alpha =      0.0500  (two-sided)
alternative p =      0.2000
sample size n =          30
```

Estimated power:

```
      power =      0.8388
```

```
. sampsi 0.05 0.2, n(30) a(.057) onesample
```

Estimated power for one-sample comparison of proportion  
to hypothesized value

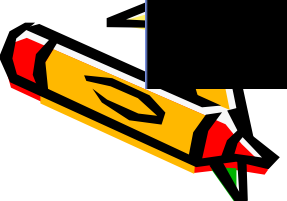
Test Ho:  $p = 0.0500$ , where  $p$  is the proportion in the population

Assumptions:

```
      alpha =      0.0570  (two-sided)
alternative p =      0.2000
sample size n =          30
```

Estimated power:

```
      power =      0.8464
```





- ある第二相試験の方法論からの抜粋。2ステージデザインを用いた。前例と同じく、5%を切ったら1st ステージせ治験中止。2<sup>nd</sup> ステージまですすんだら、20%を超えるか否かを検討することにした。
- まず15人の患者に実行し、誰も反応しなければ中止。1人以上反応すれば、更なる20人に対して試験を行う。35人中5人以上で反応がみられれば、よしとする。



- A two-stage statistical design was used to permit early termination if preliminary results indicated minimal efficacy. A target response rate of 20% was deemed sufficient to warrant further study, whereas a response rate  $\leq 5\%$  was insufficient for further investigation. This trial design therefore called for 15 assessable patients to be entered onto the first stage of the trial. If one or more responses were observed among these initial patients, an additional 20 assessable patients, would be entered. If five or more responses were observed among 35 assessable patients would be entered. If







- 本臨床試験は、10%を超えなくてはならない反応率とし、25%を目標反応率とした。サイモンのデザインによれば、50人の患者が必要で、最初の21人が6ヶ月フォローされた時点で中間解析を行う。中間解析の時点で2人以下の反応しかみられなかった場合には、その時点で試験を中止する。3人以上の反応がみられるようなら、50人のエントリーまで試験を継続する。そして、50人中8人以上で反応がみられれば、更なるステップへ進むべきであると結論する。





- This trial was designed to test the null hypothesis that the true treatment success rate is at most 0.10. The smallest treatment success proportions that would indicate that this regimen warrants further study is 0.25. The planned accrual for this **Simon-design study** was 50 assessable patients. An interim analysis was conducted after the first 21 patients had been followed for 6 months. If two or fewer responses were observed during the interim analyses, the treatment arm was to be closed permanently. If three or more confirmed responses were observed during the interim analyses, accrual was to continue. At the time of final analyses, a confirmed response among eight or more of the 50 evaluable patients would indicate that this regimen merits further investigation.



J Clin Oncol 2003; 21: 1760-6.

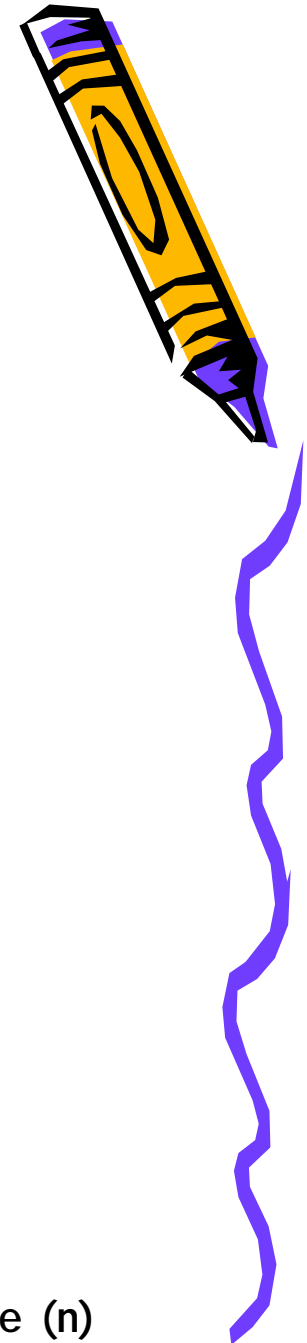




- サイモンの2ステージデザインにおける対象数計算早見表を掲載する。例えば、10%を少なくとも超え、25%を超えることを目標とする。最初の18人で2人以上に反応をみたら、次のステージにすすみ、更に25名に対して行い、7名、合計43名中11名で反応をみればよしとする。
- 真の反応率が10%のとき18人で2人しか反応しない確率は5%未満であり、真の反応率が25%のとき43人で9人が反応する確率は78%である。



	1 <sup>st</sup> stage			2 <sup>nd</sup> stage		
Reject* <sup>1</sup>	Desire* <sup>2</sup>	N1* <sup>3</sup>	R1* <sup>4</sup>	N2* <sup>5</sup>	R2* <sup>6</sup>	Max SS* <sup>7</sup>
0.05	0.20	10	0	19	3	29
	0.25	9	0	8	2	17
<b>0.1</b>	<b>0.25</b>	<b>18</b>	<b>2</b>	<b>25</b>	<b>7</b>	<b>43</b>
	0.30	10	1	19	5	29
0.2	0.35	22	5	50	19	72
	0.40	13	3	30	12	43
0.3	0.45	27	9	54	30	81
	0.50	15	5	31	18	46
0.4	0.55	26	11	58	40	84
	0.60	16	7	30	23	46
0.5	0.65	28	15	55	48	83
	0.70	15	8	28	26	43
0.6	0.75	27	17	40	46	67
	0.80	11	7	32	30	43
0.7	0.85	19	14	40	46	59
	0.90	6	4	21	22	27
0.8	0.95	9	7	20	26	29



\*1: drug not of interest if true response rate  $\leq$

\*2: desirable true response rate, \*3: accrue n1 patients

\*4: reject if  $\leq r1$  response, \*5: add n2 patients,

\*6: reject if  $\leq r$  total response, \*7: maximum sample size (n)

Simon R. Optimal two-stage designs for phase II clinical trials.  
Controlled Clin Trials 1989; 71: 1079-1985.

- 中間解析の際、反応者数が2人以下であった場合には本臨床試験を中止する。21人中2人以下の確率は7%であった。逆に3人以上で反応率が認められたため、患者リクルートは継続された。



If two or fewer responses were observed during the interim analyses, the treatment arm was to be closed permanently.



```
. bitesti 21 2 0.25
```

N	Observed k	Expected k	Assumed p	Observed p
21	2	5.25	0.25000	0.09524

→ Pr(k >= 2) = 0.980973 (one-sided test)  
Pr(k <= 2) = 0.074523 (one-sided test)  
Pr(k <= 2 or k >= 9) = 0.130670 (two-sided test)

```
. bitesti 21 3 0.25
```

N	Observed k	Expected k	Assumed p	Observed p
21	3	5.25	0.25000	0.14286

→ Pr(k >= 3) = 0.925477 (one-sided test)  
Pr(k <= 3) = 0.191682 (one-sided test)  
Pr(k <= 3 or k >= 8) = 0.321596 (two-sided test)



If three or more confirmed responses were observed during the interim analyses, accrual was to continue.

- 対象人数が最終的に50人に至ったため、解析を行う。50人中10人が反応したとすれば、20%の95%信頼区間は10-34%である。

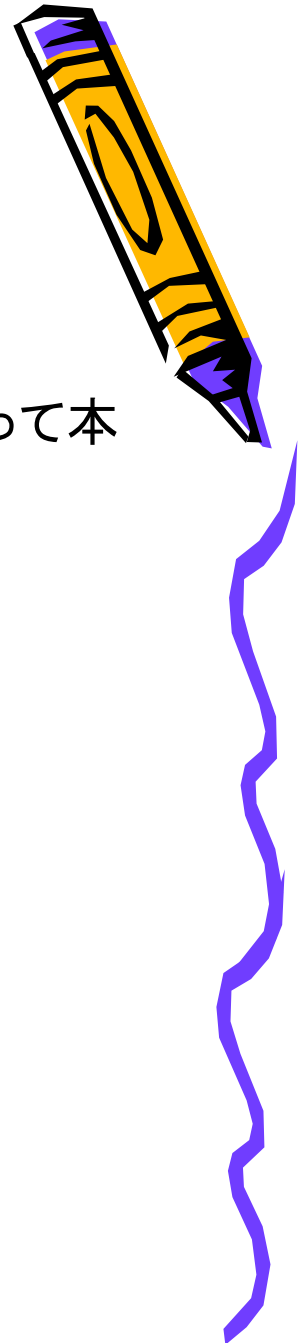


- At the time of final analyses, a confirmed response among eight or more of the 50 evaluable patients would indicate that this regimen merits further investigation.





- もしも、50人中10人以上が反応すれば、95%の信頼性をもって本治療は10%を超えていると判断できる。





```
. cii 50 8
```

Variable	Obs	Mean	Std. Err.	— Binomial Exact — [95% Conf. Interval]	
	50	.16	.0518459	.0717008	.2911263

```
. cii 50 9
```

Variable	Obs	Mean	Std. Err.	— Binomial Exact — [95% Conf. Interval]	
	50	.18	.0543323	.0857621	.3143694

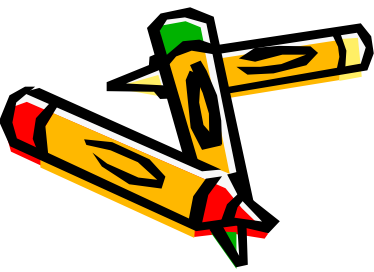
```
. cii 50 10
```

Variable	Obs	Mean	Std. Err.	— Binomial Exact — [95% Conf. Interval]	
	50	.2	.0565685	.1003022	.3371833



If tumors of 10 in 50 patients regress, true response rate will be no less than 0.10 with 95% confidence.

- 今まで、第二相臨床試験の対象数計算を行ってきた。しかし、通常の計算を用いずに単純に100例としているものもあった。それでも peer reviewed journal に掲載されている。





- Fleming single-stage procedure and tested the following  $H_0: p \leq 15\%$  and  $H_1: p \geq 30\%$ , with  $\alpha=2.5\%$  (one-sided) and a power of 90%. To allow for premature withdrawals from the study, the planned sample size was 100 patients with CML in blast crisis.
- .....; this sample size was based on **practical** considerations rather than a formal sample-size calculation.

