

Monitoring Clinical Trial

臨床試験では何故stopping rule が必要なのですか？

ACGT076 では AZT とプラシーボを比較しました。最初 900 人の HIV 陽性の妊婦がその研究を統計学的に解析する際必要であると考えられていましたが、最初の途中解析 (interim analysis) でプラシーボが 25% の垂直感染率であったのに対して、AZT 使用群では僅か 8% と大きな差がでてしまいました。その差は統計学的に有意であり、研究は中止となりました。

もし統計学的に明らかであれば、その後研究を継続しても得られる結果は同じであり、ランダム割り付けを中止して、プラシーボ群に割り付けられた妊婦に AZD を投与すれば相当数の垂直感染を防ぐことができます。しかし統計学的に明らかとは具体的にどのポイントを指すのでしょうか？もし途中解析で p が 0.05 未満であれば研究を中止するべきでしょうか？新しい患者さんのデータが加わるたびに解析を行なって、 p が 0.05 を割った時点で研究を中止するべきでしょうか？偶然 p が 0.05 を割ってしまったことも考えられるのでしょうか (false positive)？

基本的ポイント

臨床試験の間、治療効果あるいは副作用をモニターする必要がありますが、一般的にモニターは investigator と別の人が行なわなくてはなりません。そこで、通常は data safety monitoring board (DSMB) を施設内に設置します。アメリカでは 1 つの研究に対して 2 人の医師と 1 人の統計学者といった組み合わせが多いようです。研究に参加している患者さんを診療している医師や principal investigator (chair of study) は DSMB の中に含めません。何故ならバイアスを生じたり、正しく中止のタイミングを計れないかもしれないからです。

途中解析時に行なうべきこと

1. ランダムに振り分けた比較 2 群間で予後因子が偏っていないかどうかを検討します。偏りがある場合解析前に調整します。
2. 特に double blind でない研究においては、反応率を評価する際のバイアス混入の有無、あるいは程度について検討します。
3. Missing data の影響について検討します。特に 2 群間で missing data の分布に偏りがある場合にはバイアスを生じます。
4. 参加者はルールを守って治療を受けているかどうか？守っていない人はどの程度か？そのことによって研究はどの程度影響を受けるか？について検討します。
5. 可能性のある副作用と 2 番目、3 番目の優先順位で着目しているデータについても考慮します。
6. 特に多施設共同研究において、施設間で治療成績に極端な偏りがいないか検討します。もしもある場合にはバイアスが発生している可能性があり、調査すべきです。また途中で研究を中止する前に、ある施設の極端なデータに合わせたデータが強く影響を受けていないか注意深くみておく必要があります。

7. 臨床試験が長期に及ぶ場合には生存曲線解析まで行ないます。
8. 同時進行の類似の研究に注意を払います。
9. 研究中止の判断は、研究の背後に潜む目的によって異なりますが、一般的に industrial trial ではなかなか研究を中止しない傾向にありますし、academic trial では差を認めれば早期に中止する傾向にあります。

いつ、何回、途中解析を行ないますか？

一般的には半年毎に行なわれます。多施設共同研究ではそうしょっちゅう会合を開けないという事務的な理由からです。

0 ————— 6 ————— 12 ————— 18 ————— 24

先に述べたようなことを途中解析で行ないませんが、「効果判定に関しては保守的に」、「毒性に関しては自由に」、が原則です。副作用、毒性から研究中止の判断を下すときには、もっぱら経験的なものからであり、統計学的判断は介在しません。つまり、仮に1例でもきわめて重篤な問題が発生し、今後その副作用が発生すると考えられる場合は中止することもあり得るということです。DSMB が判断し IRB および PI に中止を示唆します。最終的には PI が中止を決断しますが、DSMB のコメントは強力です。

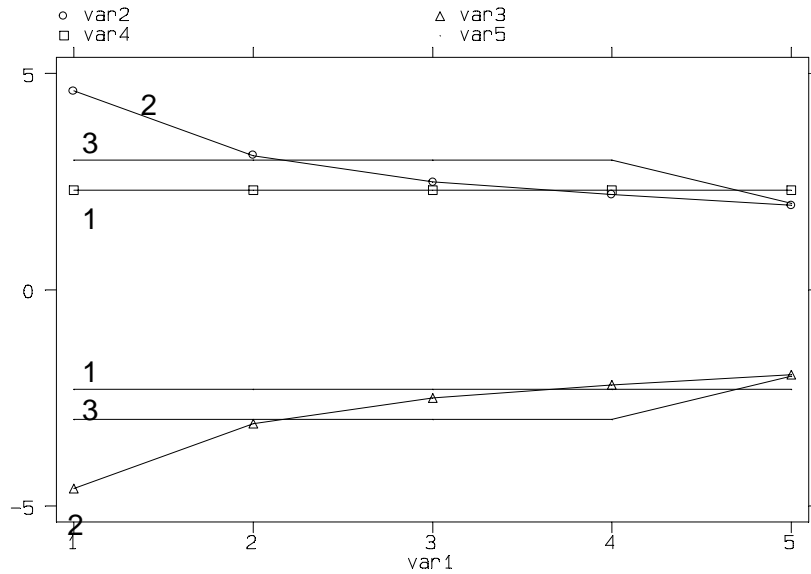
保守的効果判定とはどのような意味でしょうか？

臨時解析の数	Type I error を起こす確率 (p<0.05)
1	0.05
2	0.08
3	0.11
4	0.13
5	0.14
10	0.19
100	0.37

Type I error とは、本当はその薬が水と同じで何も効果がないのに、薬が効くと宣言してしまう確率で、頻回に途中解析を行なえば行なう程 type I error を起こす確率が高くなります。例えば5回途中解析すると、本当は効果が無いのに「効果あり」として途中で中止してしまうかもしれません。頻回に途中解析を行うことによってかえって誤った判断を下してしまう可能性があるということです。ですから途中解析は参加者1人がエンドポイントに達する毎に行うのではなく、半年毎など適当な間隔を空けて行うべきなのです。何も解析者が怠慢なわけではありません。

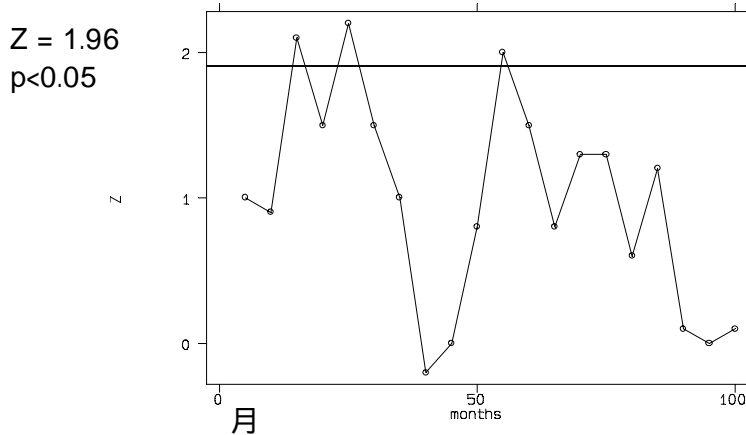
p 値がいくつ未満のとき中止しますか？

Pocock はイギリスの統計学者で、どの途中解析においても同じ p 値でよいとしています。O'Brien-Fleming は、人数が少ない時はどうしても type I error を起こしやすいので保守的な P 値、すなわち、1 回目の途中解析では z を相当大きく、そして試験が進行するにつれて cut off の z を低くしていきます。そして Peto はその中間的な立場をとっています。どれが絶対的というわけではありませんが、2 の O'Brien-Fleming の方法がよいかもしれません。



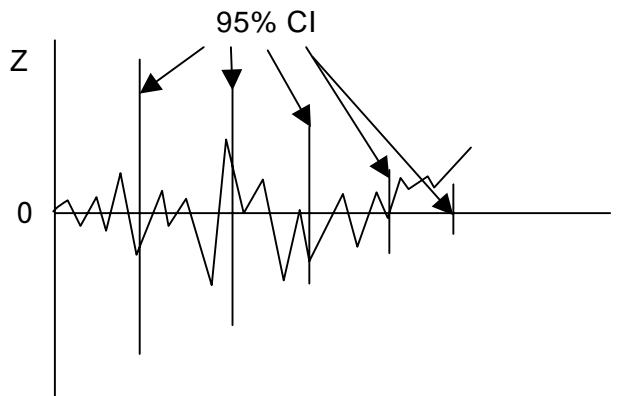
1. Pocock
2. O'Brien-Fleming
3. Peto

抗脂血症治療薬の randomized, double blind, multi-center trial が行なわれました。3 回程その薬が有効であると思われましたが ($p < 0.05$)、途中で中止せず最後まで観察したところ、結論は薬効なしでした。実際の話です。



比較的早期 Z は 1.96 を超え治療が有効かと思われましたが、続行したところ最終的には効果なしと判定された例です。対象数が少ない状態では偶然どちらかに偏りを生じることはしばしばあるのです。

P 値を用いるよりは 95%CI を用いる方が賢明かもしれません。



経過観察期間

数が少ないうちは 95%CI も広くなります。しかし人数が増えれば 95%CI も狭くなります。Z 値は同じような揺れをもって推移していますが、95%CI が小さくなることによって有意差が発生しています。

実際途中解析が予想と大幅に違う、デザインを変更せざるを得なくなることもしばしばあります。その際バイアスが入らないように注意しなくてはなりません。

例えば「全体としては有意差がないが若い人の方が薬に反応している。今後若い人を積極的に募るようにしよう。そうすれば早期に良い結果を得られるかもしれない。」と考えたら、どのような問題を発生しますか？

Randomization するので internal validity は大丈夫ですが、external validity が損なわれます。すなわち、若い人に対しては効果があると言えますが、中年、老人に対して効果があるかどうかは、その試験結果からは何も言えないことになってしまいます。

多施設共同研究

多施設共同研究における variability の最も大きな要因は施設による差です。施設によって stratification することにより影響度を評価します。以下は抗高血圧薬と placebo の 2 アームにランダムに振り分け収縮期血圧をみた臨床試験です。5 施設共同で、各施設 5 人の患者さんに投薬しています。以下のデータを左右している要因は何でしょうか？それは施設ですか、治療ですか、それとも相乗作用ですか？STATA を用いて検討してみましょう。

```
. list
```

	center	treat	BP
1.	1	1	90
2.	1	1	92
3.	1	1	93
4.	1	2	96
5.	1	2	99
6.	2	1	91
7.	2	1	89
8.	2	2	95
9.	2	2	102
10.	2	2	98
11.	3	1	92
12.	3	1	87
13.	3	1	88
14.	3	2	93
15.	3	2	98
16.	4	1	85
17.	4	1	93
18.	4	2	98
19.	4	2	100
20.	4	2	101
21.	5	1	87
22.	5	1	85
23.	5	1	92
24.	5	2	104
25.	5	2	100

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
center	25	3	1.443376	1	5
treat	25	1.48	.509902	1	2
BP	25	93.92	5.491812	85	104

```
. tabulate center treat
```

center	treat		Total
	1	2	
1	3	2	5
2	2	3	5
3	3	2	5
4	2	3	5
5	3	2	5
Total	13	12	25

. anova BP treat center treat* center

Number of obs = 25 R-squared = 0.8163
Root MSE = 2.97769 Adj R-squared = 0.7060

Source	Partial SS	df	MS	F	Prob > F
Model	590.84	9	65.6488889	7.40	0.0004
treat	493.226667	1	493.226667	55.63	0.0000
center	21.6933333	4	5.42333333	0.61	0.6607
treat*center	53.3733333	4	13.3433333	1.50	0.2506
Residual	133.00	15	8.86666667		
Total	723.84	24	30.16		

治療法の違いのみが有意な F をもっています。そなわち施設間の差は考えにくく、また施設と治療の相乗効果に関しても考えにくい状態です。