

# Randomization

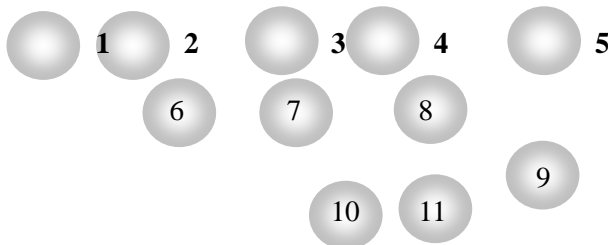
Randomization 中の imbalance をどのように解消するか？

## ランダム化の方法

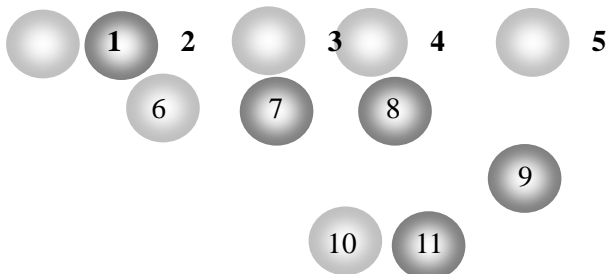
ランダム化の方法はいりいろあります。コンピュータや計算機でもランダム数についてはいます。ただ、偶数がでたら A 治療、奇数が出たら B 治療などのルールを予め決めておかなければなりません。一般的にはランダム化するセンターを作ります。ウエイ教授はかつて大学生だったころ電話番のバイトをしました。バイトの最初に「共同研究をしている 20 の病院のどこかから電話がかかってきます。ここに各病院の箱が用意してあるので、対応する箱から患者 1 人に対して 1 つの封筒を取り出して封を開け中をみて A か B を伝えてください。8 時間の電話番で 300 ドルです。」という説明を受けました。現代では、コンピュータを使用したり次ページに示すような random number table を使用します。

あるいは封筒の代わりに番号のついた薬が箱に入っている場合もあります。どの患者さんにどの番号の薬を手渡したか把握しておけば、自動的にランダム振り分けしたことになります。患者さんも手渡す方もどの薬が本当の薬でどれが偽薬かわかりません。

被験者および医師はどれが本当の薬でどれが偽薬が知りません(ダブル・ブラインド)。



モニターしている人/統計学者はどれがどれであることを把握しています。もしも、統計学者も知らなければトリプル・ブラインドです。



## ランダム表

1544	80712	97742	21500	97081	42451	50623	56071	28882	28739
01011	21285	04729	39986	73150	31548	30168	76189	56996	19210
47435	53308	40718	29050	74858	64517	93573	51-58	68501	42723
91312	75137	86274	59834	69844	19853	06917	17413	44474	86530
12775	08768	80791	16298	22934	09630	98862	39746	64623	32768
31466	43761	94872	92230	52367	13205	38634	55882	77518	36252
09300	43847	40881	51243	97810	18903	53914	31688	06220	40422
73582	13810	57784	72454	68997	72229	30340	08844	53924	89630
11092	81392	58189	22697	41063	09451	09789	00637	06450	85990
93322	98567	00116	35605	66790	52965	62877	21740	56476	49296
80134	12484	67089	08674	70753	90959	45842	59844	45214	36505
97888	31797	95037	84400	76041	96668	75920	68482	56855	97417
92612	27082	59459	69380	98654	20407	88151	56263	27126	63797
72744	45586	43279	44218	83638	05422	00995	70217	78925	39097
96256	70653	45285	26293	78305	80252	03625	40159	68760	84716
07851	47452	66742	83331	54701	06573	98169	37499	67756	68301
25594	41552	96475	56151	02089	33748	65239	89956	89559	33687
65358	15155	59374	80940	03411	94656	69440	47156	77115	99463
09402	31008	53424	21928	02198	61201	02457	87214	59750	51330
97424	90765	01634	37328	41243	33564	17884	94747	93650	77668

### 使い方

まずどのように選ぶか決めます。例えば右に2個おきに数を選ぶとします。眼をつぶって適当なところを指します。仮に一番下の一番左を選んだとしますと、9, 2, 0, 5, 6, 3, ... となります。例えばA, Bの2種類の治療であれば、偶数をA、奇数をBとします。するとB, A, A, B, A, B, となります。3種類の治療であったとすると、「1, 2, 3であればA、4, 5, 6であればB、7, 8, 9であればC、0であれば一個とばす」とすれば、C, A, B, B, となります。

### ランダム化によるimbalanceの可能性

	Difference which occurs with probability 0.05	%A / %B
10	2:8	20/80
20	6:14	30/70
50	18:34	36/64
100	40:60	40/60
200	86:114	43/57
500	228:272	46/54
1000	469:531	47/53

上の表のような偏りは5%の確率で発生します。つまり、数が少ないとランダム化しても相当偏る可能性があるということです。人数が多くなれば均等になってきます。

### Random Permuted Blocks

Imbalanceを少しでも減らす目的でブロックを作る方法があります。2種類の治療法A, Bがあるとします。

4 人を 1 ブロックとしますと

1. A, A, B, B
2. A, B, A, B
3. A, B, B, A
4. B, A, A, B
5. B, A, B, A
6. B, B, A, A

の 6 種類の組み合わせがあります。そこで先のランダム表から数を選びます。1-6 以外の数がでたときは無視します。例えば 1, 0, 4, 9, 1, 3, 0, 7, となったとすると AABB BAAB ABBB ABBA といった形で続きます。この方式をとれば全体の数が少なくても A と B はバランスよく分布します。しかし、この block design で患者さんを振り分けている人が例えば 6 を選択したとします。そして既に 2 人の患者さんを B に割り付けたとします。この臨床研究に参加している医師が「その試験が 4 文字単位の block design 方式で行われていること」を知っていれば、次の患者さん 2 人は確実に A になるとわかってしまいます。医師が次の患者さん 2 人が B になると知ったとき、バイアスを生じるかもしれません。ですから、randomization の方法論の一部は情報を収集する人に対しては伏せておかななくてはなりません。

### ランダム化でバランスがとれなかったら？

皆ランダム化すれば均等に分布すると信じていますが、上の表からもわかる通り、現実問題かならずしもそうはいきません。何故ランダム化により均等に分布するかどうかを問題視するのでしょうか？ A と B の治療法があったとすると、 $\text{variance} = \sigma^2 \{1/N_A + 1/N_B\}$  であらわされます。  $N_A = N_B$  にすると variance を最小にすることができ、より大きなパワーを得ることができます。しかし、どれくらいバランスが崩れると、どれくらいパワーが落ちるのでしょうか？

Ratio	Power
1:1	0.950
3:2	0.940
2:1	0.925
3:1	0.877

思った程落ちませんね。例えば新薬を 2、従来治療を 1 として新薬の治験を宣伝すると希望者をより多く募ることができるかもしれません。臨床試験で治療法（アーム）によりわざと比率を変えるのはそのような理由からきています。

### Stratification

それぞれの予後因子においてランダム化を行うことがあります。

	No Lymphnode meta	1-3 Lymphnode meta	>3 Lymphnode meta
A	20	40	40
B	60	25	15

例えば上の表のように、ちゃんとランダム化しても予後因子によってうまく分布していませんと、結果を解釈する際、adjustment が必要となります。

そこで上のような予後因子によって層化します。すなわち 3 つの臨床試験が同時進行するようなものです。しかし考えてみて下さい。例えば 20 の病院で多施設共同研究をや

っていて、リンパ節転移の数により3段階に分け、更に3種類の治療アームがあったとすると、180のマスを作らなくてはなりません。もし患者さんが100人しかいなければ、1つのマスは0,1,2などです。例えば小児白血病はCALLAの発現の有無、白血球の数、年齢(10歳以下、11歳以上)などが重要な予後因子ですが、これだけでも8種類のマスを作らなくてはなりません。また結果を解析するときにLogistic Regressionなどを用いて調整することもできるからstratifyする必要がないのではないかという意見もあります。しかし、全体でみると差がなくても、新しい治療が3-6歳に特に有効である場合、この重要な所見を見落としてしまいます。結局のところ患者数をできるだけ多く集められれば、それだけ多くの点について検討もできるということが出来ます。

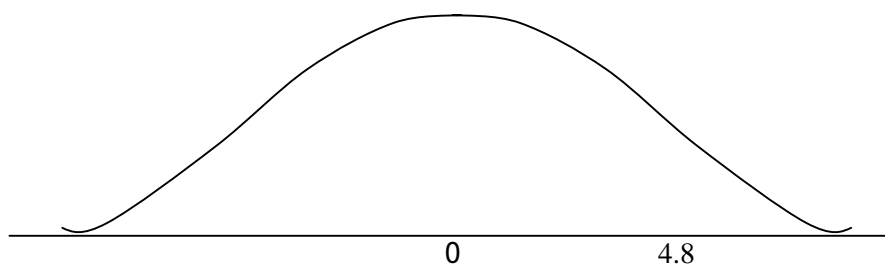
### 推論という面からみたrandomization

randomizationは何を目的に行ないますか? selection biasを減らすため、不測の因子をバランスよく振り分けるため、これらの点はよく理解されることです。しかし推論という側面からもとらえることができます。

下のような検査結果をもつ10人がいます。これをrandomizationによりa, bの2つの集団に分けました。

data	9	8	10	17	8	9	6	12	15	6
a or b	a	a	b	b	b	a	a	b	b	a

上のように5人と5人になりました。aのデータの平均は7.6、一方bの平均は12.4でありaとbの平均の差は4.8でした。全く均等に分布すれば両群の平均の差は0になるはずですが、数が少ないと3人aで7人bである、といった状況はしばしば発生します。さて、ここで10人に対するa, bの割りつけのパターンは何通りありますか? それぞれの人がa or bとなりますから、 $2^{10}$ 通りですね。そしてそれぞれの割りつけパターンに対してaの平均とbの平均の差を算出し、最後にこれらをまとめてヒストグラムを描いたとします。



これによって、たまたまランダムに得た上の表のパターンが極端な値なのか、よくあることなのか知ることができます。これをRerandomization testと呼びます。

$$\text{Var}(\text{mean}_a - \text{mean}_b) = (1/N_a - 1/N_b)\sigma^2$$

です。

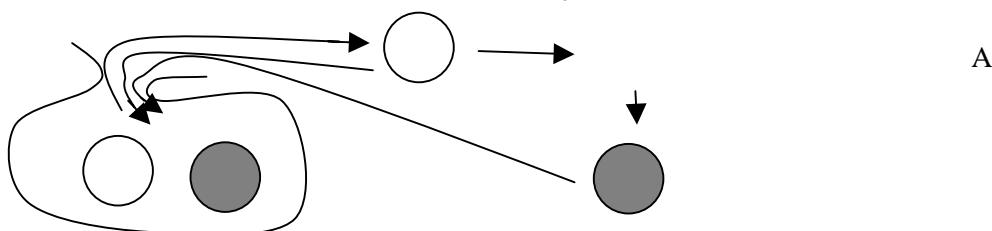
## Adaptation

我々はランダム化により偏りを生じても結果をいじることでadjustすることができます。しかし、結果的に効果がない治療に振り分けられた人は「かわいそう」です。この点何とかならないのでしょうか？特に命にかかわるような病気の場合は切実な問題です。

### Randomized play-the-winner's role (Wei & Durham JASA, 1978, Wei, Ann of Stat, 1979)

少ない数を振り分けるときしばしば偏りを生じる点は理解いただけだと思います。例えば袋に黒白の玉が5個ずつ入っています。玉を1つ取り出して、白だったらA, 黒だったらBを割り当て、またもとの袋に戻します。当然全部白で治療がAになってしまいうこともありえます。そして6個目を取り出したときBになる確率は五分五分です。

もしも最初に黒白1個ずつの玉が入っています。

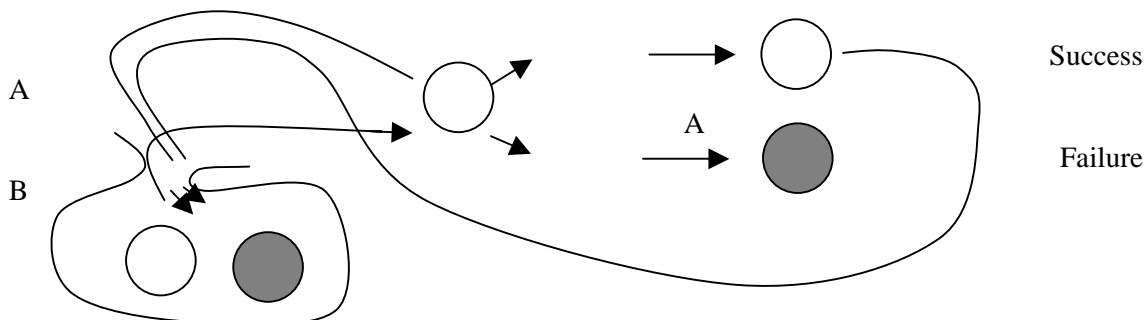


例えば、最初に白い玉を選び、Aの治療を選択したとします。そしてその白い玉を戻すとともに黒い玉を入れます。そうすると現在白い玉1個と黒い玉2個の状態です。そうすれば2個続けて白い玉がでる確率は減り、次は黒い玉の方が出やすくなります(2/3)。例えばこの方式で5個続けて白い玉が出たとします。7個目を引くとき袋の中には黒い玉6個と白い玉1個が入っており次に黒い玉を引く確率は6/7で、相当高くなります。小さなサンプル数のときよりバランスよく振り分けることができます。サンプル数が大きくなれば、通常のrandomizationに近似します。

一般的ランダム化のルールに従って参加者をA, Bに振り分け治療をはじめました。すると途中からどうやらAの方がよく効いていると思われたとします。しかも、この臨床試験の対象とした疾患が致死的な病気だったとして、あなたは統計学的有意差がでるまで継続しますか？良いと思われる治療に少しウエイトを置いて2:1の比で振り分けたりしますか？しかし、最初の時点でどちらかの治療が良いとわかっていながら臨床試験を行なう事は倫理的に問題です(行ってはいけません)。また、どちらの治療の方が良いか全く判っていないからrandomizationの意味があるのです。

### ECMO trial

1970年代新生児の肺高血圧症は人工呼吸機で治療された場合80%の死亡率でした。ミシガン大学の医師がECMO(体外循環を用いた人工肺)を用いてみようと思いました。しかし、これが有効かどうか全くわかりません。そこで彼は先のルールを用いて面白い試みをしました。



この病気に対する治療の結論は3日くらいでついてしまいます。そこで前と同様、袋に白黒2つの玉を入れておき、白が出たらAの治療、黒が出たらBの治療を選択します。それぞれ取り出した玉は再び袋に返します。そしてA治療が成功したら白い玉を追加します。すると袋の中は白2つ、黒1つですから、次の患者さんはAをひく可能性が高くなります。B治療についても同様です。この randomized play the winner's rule を適用すると、A,B治療成績の差によって患者さんの数の分布に差が生じるためパワーを失います。しかし、無用の死者を出さなくても済むかもしれません。

白い玉(A)であればECMO、黒い玉(B)であれば人工呼吸にするとして、同上の方式で患者さんを振り分けたところ以下になりました(実際の振り分け)

baby	1	2	3	4	5	6	7	8	9	10	11	12
a or b	A	B	A	A	A	A	A	A	A	A	A	A
outcome	S	F	S	S	S	S	S	S	S	S	S	S

最初の患者さんはAを引きECMO治療を受け治りました(S)。よって白い玉2個、黒い玉1個の状態です。しかし次の患者さんはBを引き従来の治療である人工呼吸により治療を受け亡くなりました。よって白い玉3個、黒い玉1個の状態です。3番目の患者さんは白い玉をひきECMO治療を受けました。その後Aがどんどん出やすくなり、結局10人の時点で比較試験を中止しました。当然ECMOに軍配があがった形です。これを Randomized play-the-winner's rule と呼びます。

	ventilation	ECMO	total
success	0	9	9
failure	1	0	1
total	1	9	10

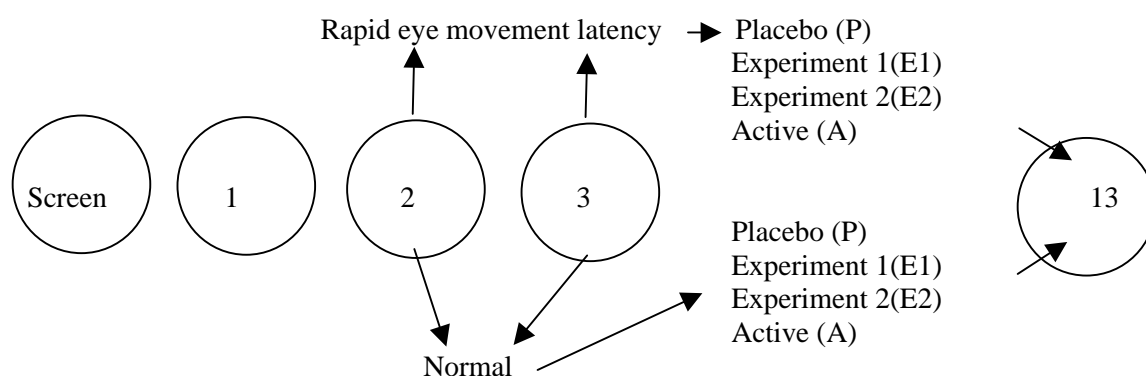
この表を Fisher's exact test で検証しても、有意差を出すことはできません。結局 2 x 2 table では、コントロールが少な過ぎて比較できないのです。比較するパワーをうしなってしまう。彼らはこれを New Engl J Med に投稿していますが、通常のランダム化が成されていないという理由で reject されています。

この結果に感化され、1985年より Harvard ECMO study が始まりました。彼らは4つずつ選ぶ block 形式でA, Bに振り分け、もしもどちらかの治療で4人の死亡がでた段階で臨床試験を中止し、勝者に軍配をあげます。そして、優っていた治療を行ない、4人死亡がでたら、あるいは28人の生存者がでた時点で治療を中止するというルールをつくり臨床試験を行ないました。最初のステージにおいて19人の患者さんが参加し、ランダム化により10人が人工呼吸、9人がECMO振り分けられました。9人のECMO治療を受けた患者さんは全員助かりましたが、従来の治療に振り分けられた10人中4人が亡くなっています。そして第二のステージでECMOを20人に試し、1人の死亡がでた時点で中止となりました。この臨床試験でもECMOの優位性が証明されたわけですが、ミシガン大学の試験より従来の治療で3人多く失っています。そして死亡した新生児の遺族は、過去に明らかな優位性を訴えた論文がありながら、臨床試験を行なったということで訴訟を起こしたのです。

この方法は、パワーを失うという弱点があるものの、「新しい治療であるが故に全く結果が予想できず、また結果が比較的早期に判るような場合」に有効です。しかし、残念ながらこの方法による割りつけは臨床試験は現実においてもあまり用いられていません。ウェイ教授らは AIDS AZT randomized clinical trial において、本法を用いた場合をシミュレーションし、大きなパワー・ロスなく(0.92 から 0.88)、76 人中 9 人の患者さんが AIDS にならずにすんだのではないかと推論しています(Yao Q, Wei LJ. Play the winner for phase II/III clinical trials. Statistics Medicine 15:2413-23,1996)。

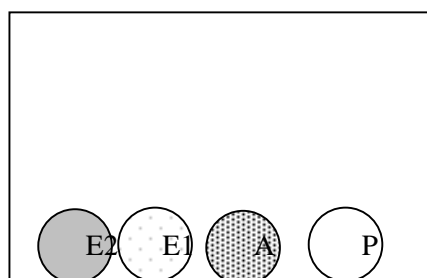
### 3 つ以上のアームがあるときのplay the winner

イーライリリーはボザックという抗うつ薬と他の新薬 2 つを比較したいと思います。まずは以下のような study design を考えました。

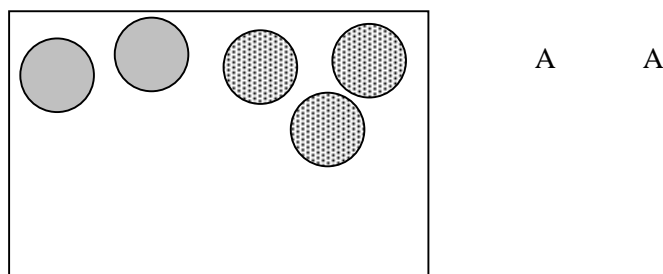


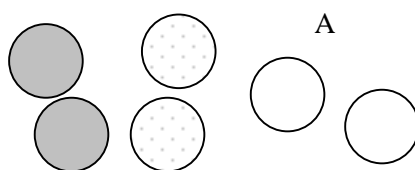
最初患者さんの状態が落ち着くまで待ちます(regression to the mean)。そして何らかの精神的刺激に対する眼の動きによって神経質群とそうでないものにスクリーニング期間に分けます。そしてそれぞれ 4 つのアームに分け 13 週後の結果を評価します。単純に考えれば、ランダムに 4 つの集団に分けて効果を比較すればよいわけですが、イーライリリーはボザックが有効なのは分かっていますから、あまりプラシーボ群の数を多くしたくありません。そこで adaptive design を考えました。

最初は 4 種類のボールが入っています。

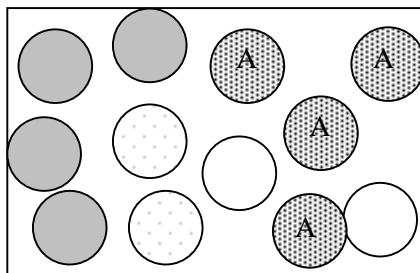


いくらか症例が集まって以下のような状態、すなわち E2: 4, A: 3, E1: 2, P: 2 となっています。

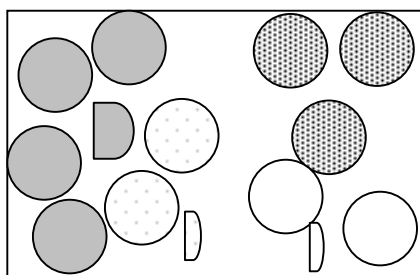




ここで次の患者さんに対して A を引き、患者さんはポザックで症状の改善をみたとします。よって A を 2 つ戻して A は 4 になります。よってその次は A を少し引きやすくなります。



さて A を引いてポザックが効かなかったとします。E1, E2, P を 1 つずつ戻すのはあまり賢くありません。現在ある数に合わせて E2 に対しては 0.5, E1 と P に対しては 0.25 ずつ戻してやります。



最終的な玉の数は以下のようにになりました。

	E1	E2	A	P
神経質	43	41	40	28
正常	64	36	53	46
合計	107	77	93	74

p の反応に対して各薬によりどのような影響がでたかを検出すればよいことになりま  
す。しかし上のような imbalance は power を落とします。しかし、その程度は思った程  
大きくありません。

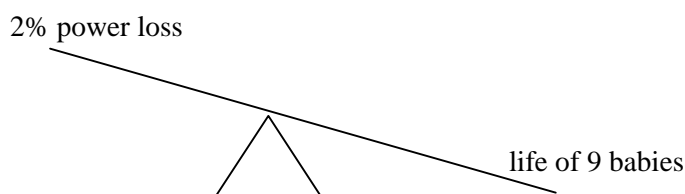
### AIDS Clinical Trial Group 076 に対する play the winner rule の効力

このプロトコールは、「HIV 陽性の母親からの垂直感染を AZT 投与によってどれくら  
い減らすことができるか試す」ものでした。476 人の HIV 陽性の妊婦が本臨床試験に参  
加し、238 人が AZT を使用、205 の出産があり、13 人のベビーが HIV 陽性となったの  
に対して(6.3%)、238 人のプラシーボ群で 204 人のベビーが生まれ、40 人が HIV 陽性とな  
りました(19.6%)。予想以上に大きな差がつき、臨床試験開始後 3 年を待たずして中  
止となっています。この時点で、もしも「全員が AZT を使用していれば 27 人が HIV  
陽性にならずに済んだ」と計算できます。

そこで play the winner に従ってシミュレーションしてみましょう。先と同様に袋に白



黒の玉が入っているところに、ある HIV 陽性妊婦が試験に参加し白を引き AZT の投与を受けたとします。そして我々は AZT の効果を知っており、7% (6.3%) の確率でベビーは HIV 陽性となります。もう1つの袋に赤い玉 7 個と黄色い玉 93 個が入っており、1つひきます。黄色い玉であれば AZT 成功ですから、白いたまを1つ追加して戻します(もとの玉と合わせると2個戻すことになる)。このように繰り返していくと AZT 群に振り分けられた妊婦は 300 であるのに対して、プラシーボは 176 になり、HIV 陽性者は 65 人となりました。元々の試験では 76 人が HIV 陽性となっていますから(中止時の人数に陽性であることが後から分かった者がいたため値が増えている)、play the winner rule で 11 人が HIV にならずに済んだこととなります。しかし、この試験では結果を得るまでに時間がかかります。よって1回ごとに玉を返すやり方は問題です。そこで8ヶ月後とか、いくつかの time period でみて少しまとめてやるようにすれば上記問題を解決できます。例えば 8, 14, 20, 26, 32, 38 ヶ月後に評価したとしますと、AZT 287 人、プラシーボ 189 人、HIV 陽性者は 67 人で 2 人増えました。この辺は trade off です。この multi-stage 法により 92% のパワーを 90% まで落としています、9 人の HIV 陽性者を出さなくて済むこととなります。



あなたはどちらに重きを置きますか？

### Dynamic treatment allocation rules

例えば 20 の施設 nioite

1,000 人の AIDS の患者さんを Protein Inhibitor (PI) の効果を調べようと思います。多くの患者さんは既に AZT を内服していますが(experience)、一部は最近診断された患者さんで AZT を内服したことがありません(Naïve)。更に HIV-RNA が 10,000 コピー以上と未満で予後が異なるので、これでも分けたいと思います。各条件に従って strata を作ると 80 になってしまい、いくつかの strata は 0 になってしまうかもしれません。そうなるので我々はパワーを失ってしまいます。

例えばある施設(#5)に HIV-RNA が 10,000 倍以上の AZT を使用したことがない新しい患者さんが臨床試験に参加することになりました。今のところの allocation は以下のような状況です。

	A (PI co-use)	B (PI no-use)
Institute #5	4	8
AZT naive	30	25
RNA > 10,000	35	27

さて、あなたなら新しい患者さんをどちらに振り分けたいと思いますか？施設#5 では B の方が多いので A にしたいところですが、全体としては AZT naïve, RNA > 10,000 の方が多いので B にしてもよいと思います。A - B の違いはそれぞれ 4, 5, 8 です。です

から絶対値としては  $RNA > 10,000$  が一番大きくなっています。しかし、1人加わることによって最も大きなインパクトがあるのはどれですか？施設#5における人数です。そう考えると施設#5としては、Aに振り分けるのが合理的といえます。

ここで、人数が少ないからAに振り分けてしまえば randomization の意味が失われず(deterministic)。そこでAの目が出やすい biased coin の原理を用います。つまりAとBの間の imbalance を測定し、stochastic の原理を用いてAに振り分けられる率を多くします。

もちろん各 strata の数が十分大きければこのような小細工をする必要はありません。しかし、各 strata の数が少ない場合には、imbalance を少なくして power を上げてやります(adaptation)。