人間栄養学講座

連載 ランダム化臨床試験をする前に

【第4回】ランダム化の方法

佐藤 俊 哉

京都大学大学院医学研究科社会健康医学系専攻医療統計学

Human Nutrition Lecture

Series : Introduction to Randomized Clinical Trials

Part 4 : Methods of Randomization

Tosiya Sato

Department of Biostatistics, Kyoto University School of Public Health

ランダム化ってなに？

臨床試験で我々が知りたいことは、「ここにいる試験参加者全員に試験的な介入を行った場合、試験参加者全員に試験的な介入を行わなかった場合と結果がどのくらい異なるか」である。ラクトフェリンのランダム化臨床試験（ラクトフェリン試験）¹⁾では、介入方法をランダムに決めることにより、我々が知りたい「試験参加者全員がラクトフェリンを摂取した場合のHCV RNA量（C型肝炎ウイルス量）の変化」と「試験参加者全員がラクトフェリンを摂取しなかった場合のHCV RNA量の変化」の比較をシミュレートすることができた。このようにランダムに分けられた2つのグループは、ラクトフェリンを摂取するかどうかは異なるだけで、それ以外のすべての要因は、年齢、性別などの直接調べることができる要因も、遺伝的素因など我々が知らないC型肝炎増悪の未知のリスク要因も、2グループ間ではほぼ均一になっていることが期待できる。

介入方法の違い以外はすべてがほぼ均一で比較可能な集団を作るランダム化という操作は、コインを投げたり、さいころを振ったり、乱数を発生させたりして実現することができるが、具体的に比較可能な2つのグループを作り出すためにはさまざまな工夫が必要となる。特に実際に臨床試験を実施する際には、試験開始時に試験参加者全員が集まって一斉に2グループに分けられるわけではなく、試験を開始してから順にインフォームドコンセントを行って、1名ずつ試験に参加

するたびにランダムに介入が決められる。本稿では、このランダム化の方法の実際について解説しよう。

単純ランダム化

通常、ランダム化臨床試験では試験介入群とコントロール介入群の試験参加者数を同数とする1:1の割り付け比が用いられる。単純ランダム化（Simple Randomization）は、この1:1の割り付け比を達成するために、2分の1の確率で試験介入Aを受けるか、コントロール介入Bを受けるかを、参加者が1名来るたびにランダムに決めていく方法である。どちらの介入を受けるかは、コインを投げて表か裏か、さいころを振って偶数か奇数かで決めることができるが、現在ではコンピュータで乱数を発生させて2分の1の確率で決めるのが普通である。コンピュータが作る乱数は、実は本当の意味での乱数ではない。コンピュータ内部の計算プログラムから産み出されるので、必ず一定の規則にもとづいている。ランダムとは予見可能でないことであつたが、コンピュータが作る乱数はその規則さえわかれば予見できてしまう。幸い、なるべく予見できないような数字の列を発生するように工夫されているので実用上問題はなく、コンピュータが発生する乱数は擬似乱数とよばれている。

単純ランダム化は非常に簡便で、わかりやすいランダム化の方法ではあるが、2分の1の確率で介入方法を決めればいつも試験介入群とコントロール介入群は

（連絡先：佐藤俊哉 〒606-8501 京都市左京区吉田近衛町 京都大学大学院医学研究科社会健康医学系専攻医療統計学
電話 075-753-4495 FAX 075-753-4487 E-mail shun@pbh.med.kyoto-u.ac.jp）

1 : 1 の同数になるだろうか。表が出る確率が2分の1のコインを10回投げるということを考えてみよう。このとき、表が5回出ることもあるし、4回のこともあるだろう。また7回表が出たり、場合によっては1回も表が出ないことも可能性は低いが起こりうることである。2分の1の確率で試験介入かコントロール介入かを決めることは、表が出る確率が2分の1のコインを投げることと同じであるから、10名の参加者を2分の1の確率で2グループに分けると、ちょうど5名ずつとなることもあれば、4名と6名、7名と3名、あるいは0名と10名に分かれることだってありうることになる。残念ながら2分の1の確率で介入方法を割り付けても、いつも2グループが同数になるわけではないことに注意してほしい。

もちろん平均的には5名ずつに分かれるのであるが、参加者数が少ないと、偶然の要素にもとづいて「9名と1名」や「0名と10名」といった極端なことも起こりうる。参加者数が増えるにしたがって、極端な食い違いはなくなっていくだろうが、それでも単純ランダム化では2グループほぼ同数になることは保証できないのである。

ブロックランダム化

いま2名の参加者、佐藤さんと鈴木さんに単純ランダム化を行って、試験介入Aかコントロール介入Bかをランダムに決めてみよう。佐藤さんと鈴木さんへの介入の割り付けは、全部で4通りあって、

佐藤さん 鈴木さん

A A

A B

B A

B B

が、すべて4分の1の確率で起こりうる。2名の参加者を単純ランダム化で試験介入かコントロール介入かに割り付けようとする、2分の1の確率で2名とも同じ介入になってしまう、こういう割り付けが起こりうるため、単純ランダム化では2グループをほぼ同数にすることはできないのである。

この4通りの割り付けで1 : 1の割り付け比にとつて都合の悪いのはAAとBBとなる場合であるので、AA、BBという割り付けはやめてしまうことにしよう。そうすると残る割り付けはAB、BAだけであるから、2名の参加者に必ず別な介入を割り付けることができる。佐藤さんと鈴木さんだけでなく、そのあとに続いて試験に参加する方たちも、2名ずつに対しこの割り付けを続けていけば、2グループの試験参加者数は最大で1名しか異ならないではないか。この割り付け方法

はもはや単純ランダム化ではなく、参加者2名ごとの組（ブロック）に介入方法をABかBAか並べ替えて（パーミューテーション）2分の1の確率でランダムに割り付けていることになるので、参加者2名を1つのブロックとしたブロックランダム化（Block Randomization）とよばれている。また、ABやBAといった介入方法の並びのことをパーミューテーションブロック（Permutation Block）という。

ブロックランダム化の問題点は、どちらの介入方法かをマスクした試験であっても、なにかの拍子に2名のうちの最初の介入群がわかってしまった場合、自動的に次の介入もわかってしまって予見できてしまうことである。しかし、いつも2名をブロックとする必要はなく、たとえば4名を1つのブロックとした場合には、2つの介入方法の並べ替えは4名から2名をとる組み合わせとなるので、全部で ${}_4C_2=6$ 通り（そうですね）

AABB ABAB ABBA

BAAB BABA BBAA

となって、4名ごとに6分の1の確率で、このうちの1つを選んで割り付ける。何名の参加者を1つのブロックとするかをブロックサイズとよび、ブロックサイズを6名、8名と増やしていくと、もし割り付けられた介入がどちらかがわかってしまっても、次の割り付けを予見できる可能性は低くなっていく。

ブロックランダム化の利点は2グループの参加者数をほぼ同数にできることにあり、予見可能性の問題もブロックサイズを大きくすることで防ぐことができる。ただし、ブロックサイズをあまり大きくしすぎると、たとえば80名の試験参加者に対してブロックサイズ24とすると、72名までは各24のブロックを全部使い切るのでAとBが同数となるが、最後のブロックを使うときにはもう8名しか残っていないので、参加者数の差が最大で8名となってしまう。参加者数をほぼ同数にするためにはブロックサイズは小さいほうがよく、予見可能性を低くするためにはブロックサイズが大きいほうがよいというジレンマになるが、通常は以下の2つの方法でこのジレンマに対処している。

1つは、1つの試験で同じブロックサイズを使い続けるのではなく、ブロックサイズ4と8からランダムに選ぶなどと、複数のブロックサイズをとりまぜて使用することである。もう1つは、割り付けに関わる担当者以外にはブロックサイズがわからないようにすることである。ランダム化臨床試験の計画書にはランダム化の方法を記載する必要があったが²⁾、試験計画書とは別に「割り付け手順書」を作成し、ブロックランダム化を用いる場合にはブロックサイズ等予見可能性に影

響するものは割り付け手順書のみに記載して、さらに割り付け手順書を閲覧できる関係者を制限する、といった措置をとっている。

層別ランダム化

ここまで「2グループの試験参加者数をほぼ同数にする」ことを前提に割り付け方法を説明してきたが、なぜ2グループの参加者数をほぼ同数にしなければならないのだろうか。1つには、一方の介入群に割り付けられる参加者数があまりにも少なくなってしまうと、ランダム化を行っていても妥当な比較ができなかったり、統計的に評価できないといったことが起こりうるからである。また、試験参加者数が前もって決まっている場合、2グループに半数ずつ割り付けると、本当に試験介入とコントロール介入の介入効果に違いがある場合、その違いがもっとも検出しやすくなるためである。

しかしランダム化でもっとも重要なことは、介入効果を検出しやすくすることではなく、あくまでも「試験参加者全員に試験的な介入を行った場合、試験参加者全員に試験的な介入を行わなかった場合と結果がどのくらい異なるか」を正しく調べることにある。このためにランダム化を行って、介入方法の違い以外の要因ができるだけ均一でバランスがとれた2グループを人為的に作り出すのであるから、各グループの参加者数が偏ることよりも、結果に強い影響を与える要因が偶然にでも一方に偏ってバランスが崩れてしまうと、介入効果が正しく調べられなくなってしまうことのほうが問題である。

性別によって結果が大きく異なることがよく知られていて、試験介入群とコントロール介入群で男女比が少しでも違ってしまうと正しく介入効果が調べられないと考えられる場合には、性別は強制的に2グループ間でバランスをとるようなランダム化の方法があると便利である。試験参加者全体の男女比が1:2であったとき、試験介入群もコントロール介入群も男女比がほぼ1:2になるように割り付けるにはどうしたらいいだろう。もちろん単純ランダム化により全員を2分の1の確率で試験介入群とコントロール介入群に割り付ければ、どちらのグループでも男女比は平均的には1:2となるはずであるが、単純ランダム化では全体の人数のバランスですらうまくとれるとは限らないので、さらに男性と女性の人数のバランスがとれるとは保証できない。単純ランダム化ではなく、ブロックランダム化を行ったらどうだろうか。ブロックランダム化は試験介入群とコントロール介入群の全体としての参加者数のバランスをとることはできるが、男女について

はやはり平均的には2グループで1:2となるものの、いつも1:2となることを保証しているわけではない。

少し見方を変えてみよう。男女比がほぼ1:2になるように2グループに分ける方法を考えるのではなく、男性は男性で1:1に2グループに、女性は女性で1:1に2グループに分けることを考えてみよう。そうすると、試験参加者全体としては男女比が1:2であったわけだが、男性は男性でほぼ半分に、女性は女性でほぼ半分に分けた2グループでも、男女比は1:2になるではないか。男性がほぼ1:1になるように2グループに分けるにはブロックランダム化を用いればよかった。女性もそうである。こうして、男女比が同じである2グループを作り出すためには、男性は男性だけで、女性は女性だけでブロックランダム化を行えばよいことがわかる。

性別に限らず、2グループ間で結果に強い影響を与える要因のバランスをとりたい場合には、試験参加者その要因で層に分け、それぞれの層の中でブロックランダム化を実施することで達成でき、これは層別ランダム化 (Stratified Randomization) とよばれている。層別ランダム化では、どうしてもバランスをとりたい要因が複数あってもよく、性別と年齢 (65歳以上と65歳未満) であれば、複数の要因を組み合わせで65歳以上男性、65歳未満男性、65歳以上女性、65歳未満女性と4つの層で層別ランダム化を実施すればよい。ただし、あまり要因が多くなると、層の数が多くなって1つの層に属する試験参加者数が少なくなってしまう。そうすると、ブロックランダム化の問題点である、最後のブロックの途中で試験が終わってしまうと参加者数のバランスが保証できなくなってしまうことから、せっかく層別したのに要因のバランスが崩れてしまうことがある。このため層別ランダム化に用いる要因は、結果に強い影響を与える要因のうちせいぜい3つまでともいわれている³⁾。

適応的ランダム化

単純ランダム化もブロックランダム化も、佐藤さんにどちらの介入が割り付けられたかと、佐藤さんの次に試験に参加した鈴木さんへの割り付けとは無関係であった。しかし、割り付けの途中で2グループ間に参加者数が大きく食い違ってしまった場合、バランスをとるため、次から参加する方たちは参加者数が少ないグループに割り付けたいこともあるだろう。このような割り付け方法は、途中の割り付け結果に適応しながら次の割り付けを決めるため、適応的ランダム化 (Adaptive Randomization) とよばれている。

単純ランダム化ではコインを投げて表か裏か、常に2

分の1の確率で介入を割り付けていたが、たとえば一方のグループに参加者数が3名以上偏った場合、参加者数の少ないグループに割り付ける確率を高くして、0.7の確率で割り付ける、というように途中で割り付け確率を変更し、2グループの参加者数の差が3名未満にもどったら再び2分の1の確率で割り付け、3名以上偏ったらまた0.7で…、という「いかさまコイン」を用いる方法が提案されている⁴⁾。

また、結果に強い影響を与える要因のバランスをとりたい場合にも、適応的ランダム化を用いることができる。性別と年齢（65歳以上、65歳未満）を考慮して割り付ける場合を例に、適応的ランダム化のなかでもよく用いられている最小化法について説明しよう。最小化法は、最初の試験参加者は2分の1の確率で割り付け、それ以降は各グループの参加者数を割り付けの際に考慮する要因で合計し、もっとも合計数の少ないグループに割り付けを行う方法である⁴⁾。表1に最小化法で性別と年齢を考慮しながら、19名まで割り付けた結果を示す。表2には、表1の結果を参加者の属性別に、性別の割り付け数と年齢別の割り付け数を合計した数をグループごとに示した。65歳以上男性では試験介入群Aが8、Bが9なので、20名目の参加者が65歳以上男性であれば試験介入群Aに割り付けることになる。同様にして20名目の参加者が65歳以上女性、65歳未満女性であってもいずれも試験介入群Aの合計数が少ないのでAに割り付けるが、65歳未満男性の場合だけA、B同数となっている。2グループの合計数が同数の場合には2分の1の確率で割り付けを行い、以下同じ手順をくり返すことになる。

適応的ランダム化にもいろいろとバリエーションがあり、表2の2グループ間の合計数の差が一定数（たとえば3）以内ならば2分の1の確率で割り付け、一定数を越えた場合に合計数が最小となるグループに割り付ける方法や、合計数の少ないグループに強制的に割り付けるのではなく、合計数の少ないグループに割り付ける確率を高くする方法などが提案されている。最小化法をはじめとする適応的ランダム化は層別ランダム化のように複数の要因の組み合わせで考えるのではなく、それぞれの要因ごとの合計数でバランスを考えると、割り付けの際に考慮する要因の数が増えても層別ランダム化のような問題は起こらないという利点がある。

クラスターランダム化

これまで解説したランダム化の方法は、すべて試験参加者1人ひとりに介入を割り付ける方法であった。一般に臨床試験では、医薬品の効果を調べる場合のよ

表1 最小化法で19名まで割り付けた結果

		A	B
性別	男性	3	3
	女性	6	7
年齢	65歳以上	5	6
	65歳未満	4	4

表2 19名までの属性性別、性・年齢の合計数

		A	B
65歳以上	男性	8	9
65歳未満	男性	7	7
65歳以上	女性	11	13
65歳未満	女性	10	11

うに、介入を行う単位が個人であるため、通常は個人単位の割り付け方法を用いることが多いが、介入方法自体が個人ではなくグループを対象としている場合がある。たとえばマスメディアを使った健康教育プログラムの効果を評価したい場合、マスメディアで健康教育プログラムを配信した場合と実施しなかった場合の比較が目的となる。

このような場合には個人にその健康教育プログラムを実施しても、目的としているマスメディアによる健康教育プログラムの効果は調べることができないので、個人単位ではなくマスメディアがカバーする地域やグループを1つの単位として、健康教育プログラムを配信するかどうかをランダムに決めることになる。介入方法の割り付けを行う単位が個人ではなく、地域やグループといった塊り（クラスター）を単位として行うランダム化のことをクラスターランダム化（Cluster Randomization）という⁵⁾。マスメディアを使った健康教育プログラムでは、目的自体からしてクラスター単位でしか効果を調べることができないが、そのほかにも高等学校での禁煙教育の効果を調べたい場合も、高校生1人ひとりに禁煙教育を行うことが目的ではなく、学級単位で禁煙教育を行った場合の効果を調べたいのであるから、クラスを単位とした割り付けを行う必要がある。また、食事療法の効果を調べたい場合のように、個人単位の割り付けで効果を調べられる介入であっても、家族単位で割り付けたほうが、1人だけ特別なメニューを作らなくてもすむようなクラスター割り付けもある。

クラスターランダム化では、個人単位でランダム化を行う試験での「個人」と「クラスター」が対応しているので、原理的には個人単位のランダム化臨床試験で必要な参加者数と同じ数のクラスター数が必要となる。これは、同じクラスターに属する人、家族やクラ

スマート、はいろいろな意味で似た特徴をもっているはずであり、そのため介入による結果も似た結果となりやすい傾向にあるためである。クラスター内で参加者の結果が似た傾向となることをクラスター内相関を使って表そう。もし同じクラスターに属する人がみな同じ結果になるのだとすると、クラスター内相関は1であり、クラスター内の1名を対象としても全員を対象としても同じ結果しかえられないので、個人割り付けの試験の必要参加者数と必要クラスター数は同じとなり、クラスターランダム化試験の総参加者数は個人割り付け試験に比べかなり多くなってしまう。

一方、同じクラスター内の参加者であっても結果がバラバラでまったく予測できない場合はクラスター内相関はゼロであり、この場合に限ってはクラスターランダム化試験に必要な参加者数と個人割り付け試験に必要な参加者数は等しくなる。現実にはクラスター内相関は0と1の間にあるはずである。クラスター単位でしか効果を調べることができない介入の場合はクラスターランダム化試験を実施せざるをえないが、個人割り付け、クラスター割り付けどちらでも実施可能な介入の場合は、クラスター内相関を考慮して、クラスター割り付けにより複数の参加者に同じ介入をいっぺんに実施できることのメリットと、それによって参加者数が多く必要となることのデメリットのバランスを考えてデザインを選ぶ必要がある。クラスターランダム化は割り付けはクラスター単位であるが、残念ながらインフォームドコンセントはクラスターの代表者ではなく、原則として個人単位で実施する必要があるので注意してほしい。

*

ラクトフェリン試験でのラクトフェリンとプラセボの割り付けは、「施設を層別因子とするパーミューテーションブロック法」を用いた¹⁾が、これはC型肝炎の患者の状態が施設によって異なることが考えられ、そのことがHCV RNA量の変化に影響することを考慮したためである。そしてラクトフェリン試験では、試験計画書とは別に割り付け手順書を作成し、ブロックサイズなどの情報は、試験統計家である私と割り付け責任者（臨床試験支援会社に委託）のみが閲覧できることにして、予見可能性を減らす工夫をした。

本稿では「試験を開始してから順にインフォームドコンセントを行って、1名ずつ試験に参加するたびにランダムに介入が決められる」状況を想定してランダム化の方法を解説したが、査読者から、栄養指導などの介入効果を調べるために、ボランティアを募り同意した方たちのリストができた時点で割り付けをするといった場合もありうるとの指摘を受けた。もっともな状

況であり、この場合のランダム化の方法について考察したい。

まず、ボランティアとして試験参加に同意した方たちのリストができた時点であっても、本稿で解説した1名ずつ介入をランダムに決める方法は適応可能となるし、リストができるまで割り付けを待つ必要もない。全試験参加者のリストがランダム化を行うときにあるメリットとして考えられることは以下の点である。

たとえば栄養指導の介入効果を調べる試験に100名の参加者が必要で、ボランティアを募ったところ100名が同意したとしよう。単純ランダム化にしても、ブロックランダム化にしても、適応的ランダム化にしても、この100名をちょうど50名ずつの2グループに必ず分けることはできなかったが、参加者が100名だということがわかっているのので、ブロックサイズが100のパーミューテーションブロックを用意しよう。

このブロックサイズが100で、半分の50が新しい栄養指導介入群、残りの50が従来指導介入群というブロックを1つランダムに選んで参加者に介入方法を割り付けることで2グループを必ず50名ずつにすることができる。実際にはブロックサイズが100のパーミューテーションブロックの総数は10の29乗となり用意することはできないが、100個の一樣乱数をコンピュータで発生し、小さい順に並べ替えて上から50番目までを新しい栄養指導介入群、51番目以降を従来指導介入群とすればよいので簡単に実施できる⁴⁾。結果に影響を与える重要な要因があらかじめわかっているならば、その要因で層別した参加者をちょうど半分ずつに分けることも可能となる。

このような場合に、単純ランダム化を行って性別や年齢などいろいろな要因を2グループで比較したら差があったからといって、もう一度ランダム化をやり直して、いろいろな要因の差がみられなくなるまでランダム化をくり返す、ということはすべきではない。ランダム化の意義は偶然の要素にもとづいて2つのグループに分けることであるのに、誰かが割り付け結果をみて、自分に都合のよい割り付けを選んでしまうと、それはもはやランダム化ではない。ランダム化は、介入効果を検定や推定といった統計的方法で調べる場合の基礎を与えているが、何回もランダム化を行って都合のよいものを選んでしまうと、検定結果や推定結果も信用できないものになってしまう。

偶然の要素にもとづいて2つのグループに分けるということの意味は、2グループ間で性別、年齢といった100個の背景要因について差がないかどうかをアルファレベル5%で検定すれば、平均して5個は統計的に有意な結果になる、ということである。5%水準の検定を

行って2グループで100個の背景要因が1つも有意とならなかったら、それはおかしいランダム化をしていることになり、試験の結果としてえられる介入効果の検定、推定にもおかしい影響を与えかねない。

ランダム化臨床試験が成功するかどうかは、ランダムに分けられた2グループ間で結果に強い影響を与える要因のバランスが十分にとれているかどうかにかかっているのは間違いないが、そのためにはこれまでの研究結果や文献のレビューから、計画している試験のエンドポイントに強い影響を与えることが知られている要因を同定しなければならないし、場当たりのではなく、その要因をきちんと考慮したブロックランダム化や適応的ランダム化を行う必要がある。

文 献

- 1) Ueno, H., Sato, T., Yamamoto, S., Tanaka, K., Ohkawa, K., Takagi, H., Yokosuka, O., Furuse, J., Saito, H., Sawaki, A., Kasugai, H., Osaki, Y., Fujiyama, S., Sato, K., Wakabayashi, K. and Okusaka, T.: Randomized, double-blind, placebo-controlled trial of bovine lactoferrin in patients with chronic hepatitis C, *Cancer Science*, **97**, 1105–1110 (2006)
- 2) 佐藤俊哉：ランダム化臨床試験をする前に【第2回】試験計画書の作成, 栄養学雑誌, **65**, 149–153 (2007)
- 3) Green, S., Benedetti, J. and Crowley, J.: *Clinical Trials in Oncology*, 2nd ed. (2003) / 福田治彦, 新美三由紀, 石塚直樹訳: 米国 SWOG に学ぶがん臨床試験の実践, pp.56–58 (2004) 医学書院, 東京
- 4) 西 次男: 6 無作為化, 臨床試験ハンドブック / 丹後俊郎, 上坂浩之編, pp.137–155 (2007) 朝倉書店, 東京
- 5) Murray, D. M.: *Design and Analysis of Group-Randomized Trials* (1998) Oxford University Press, New York