

# 第8章

## 生物統計学2: 交絡・ランダム化と因果推論





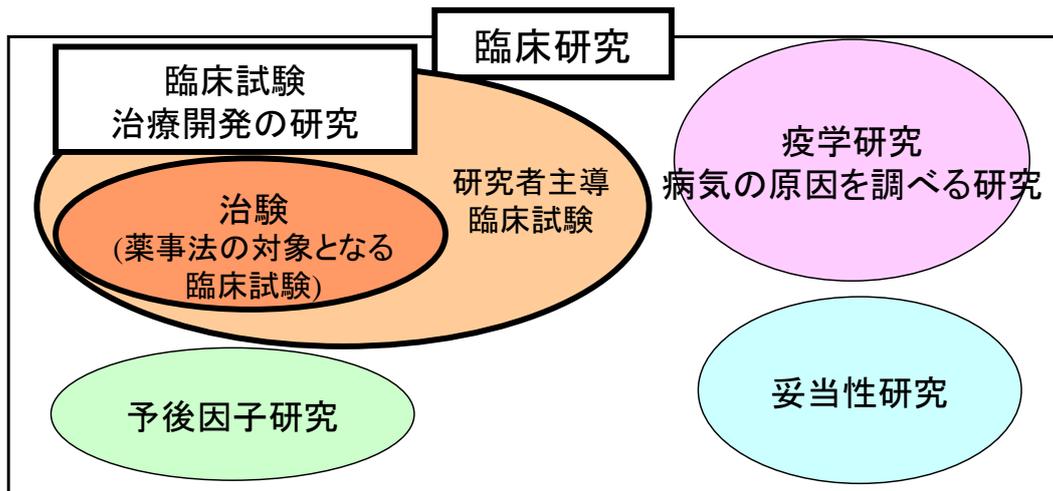
# 本講義の内容

- 交絡とその調整
- 因果推論とランダム化



本講義で説明する交絡とその調整、因果推論とランダム化は、臨床研究を支える生物統計学において重要な概念といえます。

# 臨床研究とそれを支えるもの



生物統計は、臨床研究分野で用いられる統計学のこと



## 生物統計学とは？

視聴者4人に毎日眠る前、ミニトマトを3つ摂取してもらう



一週間後、被験者4人中3人も血液がさらさらに！



生物統計学とは、臨床研究領域で用いられる統計学を指すことを前講義でご紹介しました。

ここでは、仮想的な健康情報TVショーで行われたある実験結果を例に、臨床研究を理解する際に必要な生物統計学の基本的考え方について説明します。

ある健康情報TVショーでは、視聴者4人に毎日眠る前にミニトマトを3つ食べてもらう実験を行い、1週間後、被験者(実験に参加した人)4人中3人(75%)の血液がさらさらになったという結果が提示されました。

「ミニトマト3つで血液がさらさらになる」ことが本当かどうか、このショーを見て判断できるでしょうか。

# 結果は解釈可能か？

さらさら割合 75%



**3/4 = 75% に対して効果** が得られた

- 4人中3人ぐらいは偶然でも起こる**かもしれない**  $\alpha$ エラーの問題
  - 4人の結果では少なすぎて信用できない**かもしれない**  
サンプルサイズ、検出力、 $\beta$ エラーの問題
  - (他にも血液さらさら効果があるといわれる)  
黒豆とか、納豆とか、青汁とか、、、摂取した**かもしれない**
  - ミニトマトを食べなくてもさらさらになった**かもしれない** 交絡の問題
- 様々な「**かもしれない**」により結果が適切に解釈できない  
統計学を用いれば適切に対処できる



Introduction to Clinical Research

5

75%にさらさら効果が得られたという結果に対して、

「4人中3人ぐらいは偶然でも起こるかもしれない」

「4人の結果では少なすぎて信用できないかもしれない」

「ほかのさらさら効果があるものを摂取したかもしれない」

などが疑問点として挙げられます。つまり、75%に効果があるという結果を得ましたが、そこにはいくつもの「かもしれない」原因が潜んでいるため、適切に結果を解釈できない状況といえます。

統計的により厳密にこれらを整理していくと、1つ目の「かもしれない」は $\alpha$ エラーの問題と整理できます。

また2つ目の「かもしれない」はサンプルサイズ・検出力・ $\beta$ エラーの問題、

残り2つの「かもしれない」は交絡の問題と整理できます。

# 生物統計学とは？



## こういう場面で 生物統計学の考え方は非常に重要

- 生物統計学の考え方を
  - データ測定前の研究計画段階の場面に
  - 測定後のデータ解析の場面に用いれば様々な「**かもしれない**」が存在することによって結果が解釈できなくなってしまう可能性を可能な限り小さくできる
  - どちらかといえば研究計画がより重要
    - 生物統計家の技量がより試される



このように、結果を適切に解釈できないような場面で、生物統計学の考え方は非常に重要となります。特に「研究計画段階」、「データ解析」の2つが、生物統計学の考え方が重要となる場面です。

## 他にも「かもしれない」



3/4 = 75% に対して効果

- 4人中3人ぐらいは偶然でも起こる**かもしれない** ✓  $\alpha$ エラーの問題
- 4人の結果では少なすぎて信用できない**かもしれない** ✓  
サンプルサイズ、検出力、 $\beta$ エラーの問題
- ミニトマトを食べなくても、さらさらになった**かもしれない**
- 他にも(血液さらさら効果があるといわれている)  
黒豆とか、納豆とか、青汁とか、、、摂取した**かもしれない**



を食べなかった人と比較しよう！



4つのかもしれないのうち2つ目までは、前講義で説明した  $\alpha$  エラーとサンプルサイズ、検出力、 $\beta$  エラー の問題であると整理できました。

しかし、まだ2つ問題が残されたままです。

この問題を解決するために有用な方法は、「ミニトマトを食べなかった人と比較する」ことです。

# 「比較をしよう！」

- 視聴者50人に毎日眠る前、ミニトマト3つ**摂取してもらって**一週間後に血液さらさらか調べる
- それとは別に、毎日眠る前にミニトマトを**摂取しなかった**視聴者50人を募集して、一週間後に血液さらさらか調べる

以下のように仮説を変更して比較

帰無仮説「さらさら割合の差無し(差が0%)」

対立仮説「さらさら割合の差30%」

$\alpha = 5\%$ 、検出力約**85%**で必要サンプルサイズ計**100人**



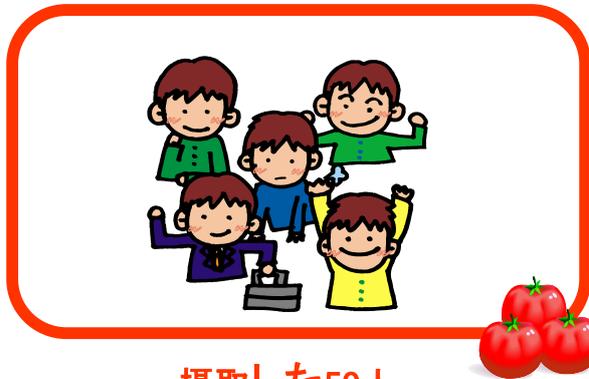
ミニトマトの効果を調べるために比較をすることを考えます。

ここでは、ミニトマトを摂取した群の50人と摂取しなかった群の50人を比較するという実験方法を採用し、帰無仮説、対立仮説、 $\alpha$  エラーの大きさ、検出力などをあらかじめ決め、実験します。

計100人の場合、設定した対立仮説に対する検出力は85%となります。

## 2つの群の比較

### • 実験デザイン



摂取した50人

### コントロール (対照)



摂取しなかった50人

- 一週間後、2つのグループの結果  
(血液がさらさらになった視聴者の割合)を比較



実験デザインは、摂取した50人と摂取しなかった50人との比較となります。  
ここでは「ミニトマトを摂取した群」で血液がさらさらになる効果があるかどうかに興味があります。この興味のある群の比較相手とする群のことを「コントロール(対照)」といい、ここでは「ミニトマトを摂取しない群」がこれに該当します。

## 割合の差(リスク差)は？

	さらさら	効果なし	
 食べた	35人(70%)	15人	50人
 食べず	25人(50%)	25人	50人

$$70\% - 50\% = 20\% \text{の差}$$



を食べたグループで20%さらさら割合が高い

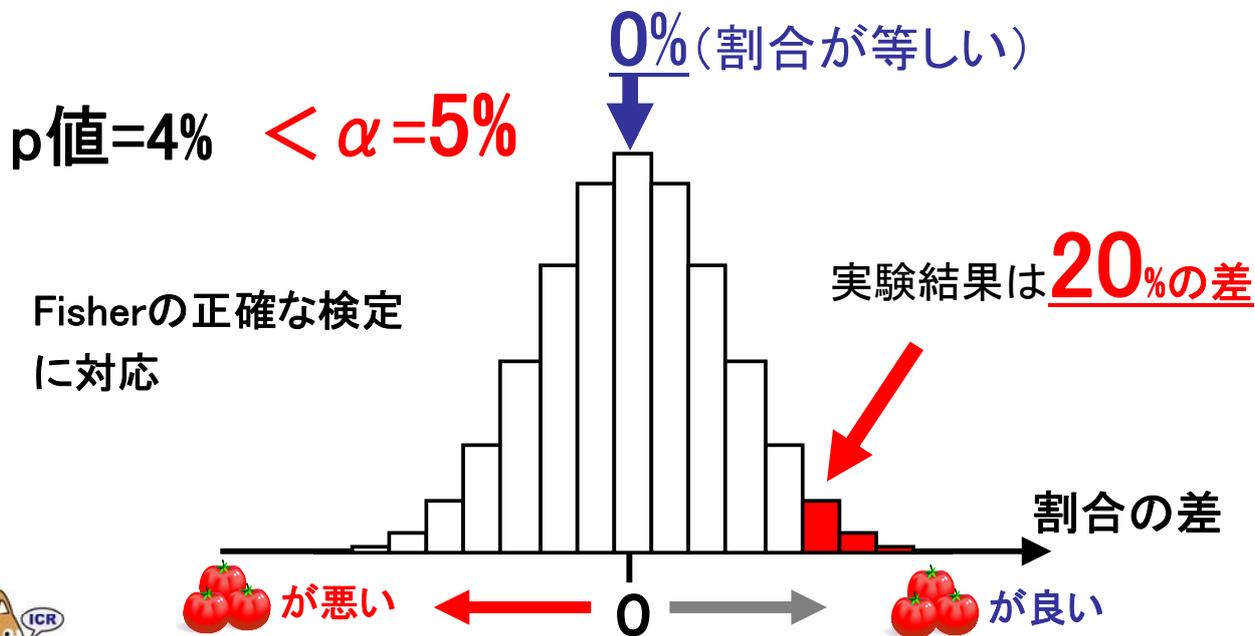


実験の結果、ミニトマトを摂取した50人のうち、血液がさらさらとなった人が35人(70%)、さらさらとならなかった人(効果なし)が15人となりました。一方、コントロールである摂取しなかった50人では、さらさらとなった人とさらさらとならなかった人はそれぞれ同数の25人(50%)でした。

割合の差は $70\% - 50\% = 20\%$ と計算できます。摂取した方が20%さらさらとなる割合が高いことを表します。割合の差を生物統計学や疫学では特にリスク差と呼ぶことがあります。

# 前回と同様に p値 を求める！

- 帰無仮説「さらさら割合が等しい」が正しい場合  
結果が偶然でどの程度ばらつくか



この結果から、前回と同様に仮説検定を行い、p値を求めると4%となります。

今回、事前に  $\alpha = 5\%$  と判断規準を定め、事前に定めた規準よりも小さな値としてp値が観察されたので、統計的有意差ありと判断できます。

これにより帰無仮説を却下することになります。

## この結果をみて

- ミニトマトの効果はある！ 統計的有意差！
  - 食べなかったときと食べなかったときのさらさら割合の差は20%もある ( $p=4\% < \alpha$ )
- 名物司会者は「リコピン・パワー！でさらさら」と得意げに生物学的機構を説明しながらオススメ
  - 夕方のスーパーマーケットでは、ミニトマトが完売
- みなさんはこの結果から  を買いますか？  
今日から買って食べます or これだけやっても信じません



では、この比較の結果をみて、ミニトマトに血液さらさら効果があると判断できるでしょうか。

## ところで

- 違う曜日のライバルTV健康ショーが
  - ブロッコリの大腸がんに対する予防効果をうたっていた
- もしかして血液さらさら効果もあるかも(新しい仮説)
  - 生物学的機構: ブロッコリに含まれるイソチオシアネートの効果?
  - 先ほどのミニトマト実験の100人について



を食べていたかどうか聞いてみよう



血液さらさら効果があるかもしれないものとして、ミニトマトの他にも、ブロッコリがあります。ブロッコリに関しても調べてみることにしましょう。



## ブロッコリの効果

- ブロッコリを食べた人と食べなかった人とで比べよう

	さらさら	効果なし	
 食べた	60	0	60人
 食べず	0	40	40人

割合の差は  $100 - 0 = 100\%$ 、めちゃめちゃ効果あり



ブロッコリでも同様に、食べた人と食べなかった人とを比べてみると、ブロッコリを食べた場合の、食べなかった場合に比べた割合の差は100%となり、ミニトマトよりも大きな割合の差が得られたとします。



## ブロッコリも効果ありということ？

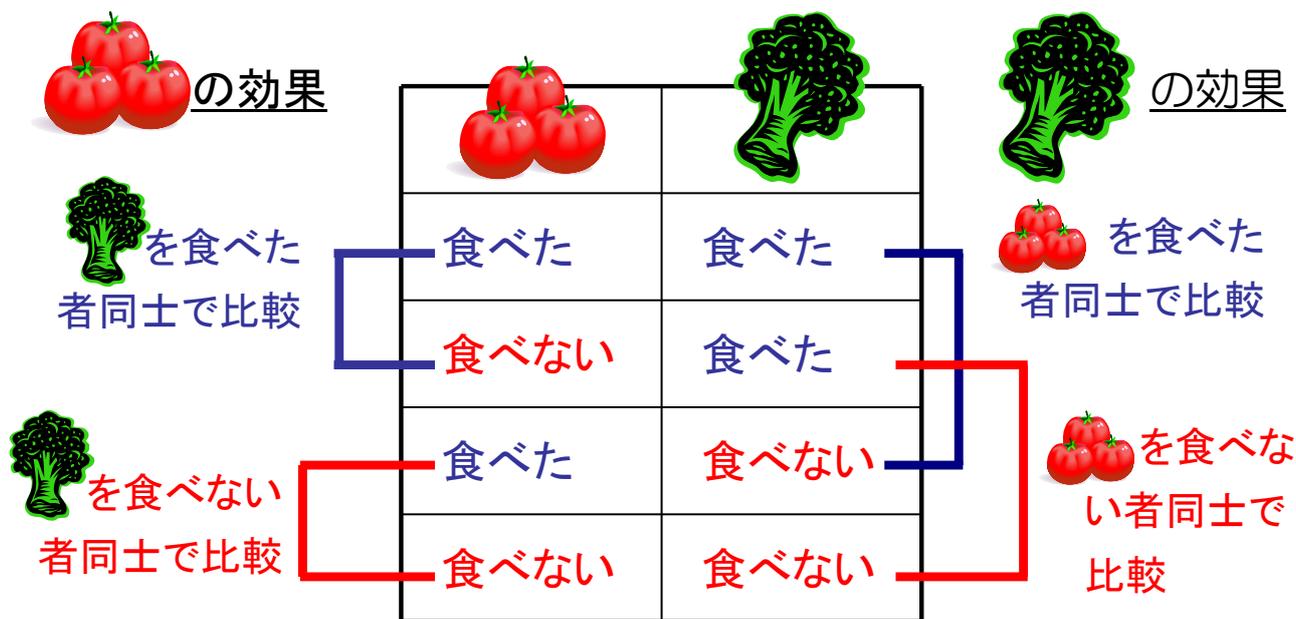
- ブロッコリのさらさら効果は絶大
- そうすると、どういう解釈ができるか…
  -  と  共にさらさら効果あり
  -  のみにさらさら効果あり
  -  のみにさらさら効果あり
- 個別の結果だけではどれが正しいか判断できない！
  - 本当はどちらに効果があるのか分からない



このような結果が得られた場合、それぞれどのように解釈することができるでしょうか。

個別に評価した結果だけでは、それが正しいかどうかは判断できないので、ミニトマトとブロッコリを一緒にみてみることにします。

# サブグループ解析(層別解析)の考え方



サブグループごとに比較すれば互いの影響を取り除くこと(調整すること)ができる



ミニトマトとブロッコリ、それぞれに関して適切な判断を導くよい方法として、サブグループ解析(層別解析)があります。

サブグループ解析では、例えばミニトマトに興味がある場合、表の左に示すようにブロッコリを食べた者同士、或いはブロッコリを食べない者同士で比較を行います。

また、ブロッコリに興味がある場合にも、表の右に示すように同様に比較することができます。

ここでの要点は、興味があるもの以外の条件を等しくすることにあります。

# ミニトマトで分けて の効果を見ると



食べた人のみ

		さらさら	
		さらさら	効果なし
 食べた		<b>35</b>	<b>0</b>
	35		
 食べず		<b>0</b>	<b>15</b>
	15		



食べなかった人のみ

		さらさら	
		さらさら	効果なし
 食べた		<b>25</b>	<b>0</b>
	25		
 食べず		<b>0</b>	<b>25</b>
	25		



を食べた人でも食べなかった人でも



のさらさら効果は絶大！



それでは、実際にサブグループ解析をしてみます。

まずは、ミニトマトで分けてサブグループ解析を行い、ブロッコリの効果を見てみることにします。そうすると、表に示す結果が得られました。

ここから、ミニトマトを調整したとしても、ブロッコリの血液さらさら効果は絶大であると判断できます。

# ブロッコリで分けて の効果を見ると



食べた人のみ

		さらさら	効果なし	
 食べた		<b>35</b>	<b>0</b>	35
 食べず		<b>25</b>	<b>0</b>	25



食べなかった人のみ

		さらさら	効果なし	
 食べた		<b>0</b>	<b>15</b>	15
 食べず		<b>0</b>	<b>25</b>	25



を食べた人でも食べなかった人でも



のさらさら効果はゼロ！



反対に、ブロッコリで分けてサブグループ解析を行ってみると、どちらにおいてもミニトマトの効果はありませんでした。

# どうしてこのようなことになったのか

• 人数をよくみると



を食べた人に



を食べた人が多い

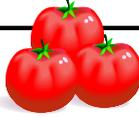
• サラダと一緒に摂取？



の影響により見かけ上



の効果があるようにみえてしまった

		人数
食べた	食べた	<u>35人</u>
食べた	食べない	∨
食べない	食べた	<u>25人</u>
食べない	食べない	



どうしてこのようなことになったのかというと、ミニトマトを食べた人にブロッコリを食べていた人が多かったからです。

# この現象が交絡 confounding

- ラテン語で“一緒に混ぜる” confundere が語源
- 結果に影響を与える**他の要因**との関連により、**ある要因**の結果への影響が見かけ上歪められてしまうこと

## 今回の場合



- 結果に影響を与える**他の要因**との関連により、**ある要因**の結果への影響が見かけ上歪められてしまうこと
-  を**交絡因子**という



この現象は、疫学や統計学上、非常に重要な概念であり、「交絡 confounding」と呼ばれます。ここではブロッコリとの関連によって交絡が生じ、これによりミニトマトの結果への影響が歪められてしまっていました。

ここでいう「ブロッコリとの関連」とは、ミニトマトを食べた人の中にブロッコリを食べた人が多かったことを表します。

また、ブロッコリのように交絡が生じる原因となる因子のことを、「交絡因子」と呼びます。

# 交絡因子 = 調整しなければならない変数

- ① 興味のある曝露因子(  )と関連する
- ② 結果( さらさら )に影響する
- ③ 興味のある曝露因子(  )の結果でない

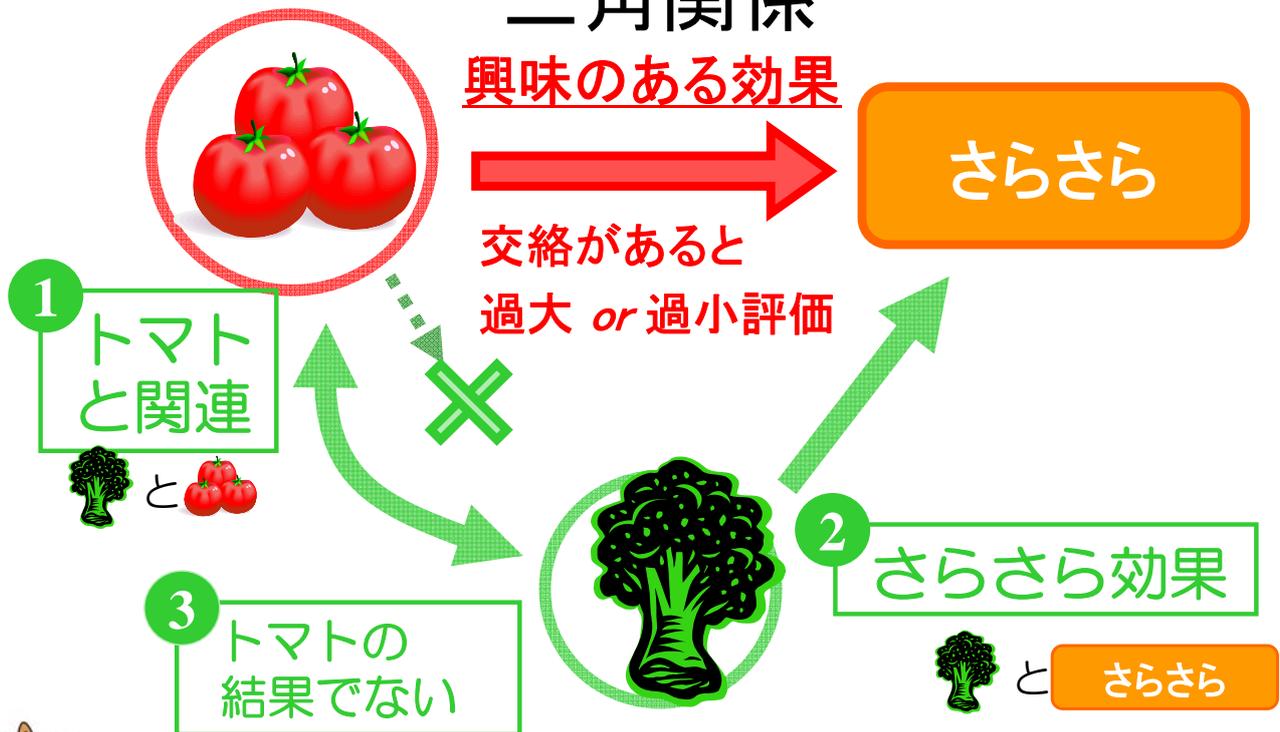
- 3条件全てを満たす場合に交絡因子、交絡を生む
- 交絡があると、結果を適切に解釈できない
  - 過大評価(効果がないのに効果あり)にも、  
過小評価( あるのに なし)にもつながる
  - 適切に調整しないと解釈できない



ある因子が交絡因子となるためには、①～③の3つの条件を全て満たす必要があります。

今回の場合、ブロッコリは3条件を全て満たしているため交絡因子となっていました。

# 交絡因子の3条件 ①~③ と交絡の三角関係



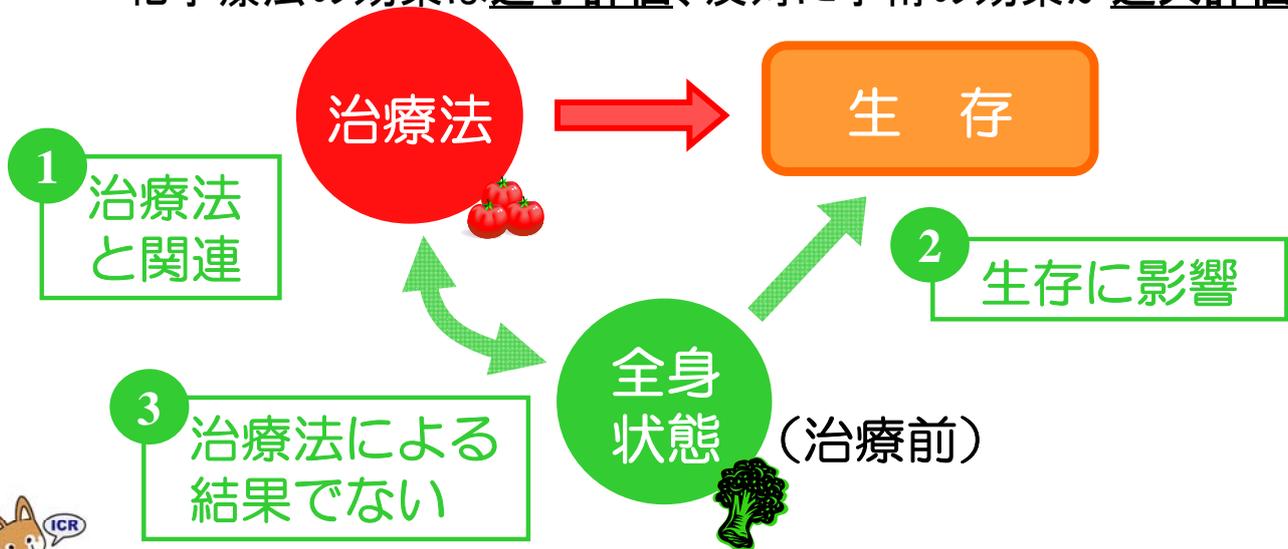
この図では、交絡因子の3条件がブロッコリーの例でもあてはまるかどうかを確認していきます。

ブロッコリーは、

- ① トマトと関連し、かつ
  - ② さらさら効果を持ち、かつ
  - ③ トマトの結果ではないため、
- 交絡因子の条件を全て満たしていることが分かります。

## 他にも、、、

- 「手術の方が化学療法に比べて予後がいいんだよね」
  - 手術が受けられないほど治療前の全身状態が悪い患者さんのみが選択的に化学療法を受けているのであれば  
化学療法の効果は過小評価、反対に手術の効果が過大評価



医学領域でよく挙げられる交絡因子の例としては、治療法評価の際の全身状態などがあります。

# 適切に交絡調整して初めて解釈できる



食べた人のみ

	さらさら	効果なし
 食べた	<b>35</b>	<b>0</b>
 食べず	<b>25</b>	<b>0</b>



食べなかった人のみ

	さらさら	効果なし
 食べた	<b>0</b>	<b>15</b>
 食べず	<b>0</b>	<b>25</b>



を食べた人でも食べなかった人でも



のさらさら効果は完全にゼロ!



ミニトマトとブロッコリの実験は、適切に交絡因子を調整してみて初めて、結果が解釈できる例であったことが分かります。

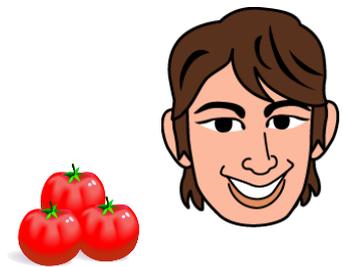
# 被験者の中田くんを考えてみよう！

## 実験開始日

- 激戦続きでお疲れの「中田くん」
- 就寝前に  を食べた

## 1週間後

- 血液がさらさらになった
-  はさらさら効果があった



ところで、そもそもどのような状態を「効果あり」、「効果なし」というのでしょうか。

中田くんにミニトマトを食べる被験者となってもらい、思考実験を試みることにしましょう。

1週間後、中田くんの血液はさらさらになり、ミニトマトは「効果あり」と判断しました。

しかしながら、本当に効果はあったのでしょうか。

# でも、もしかしたら

**反事実** counterfactual  
を考える

## 実験開始日

- 激戦続きでお疲れの「中田くん」
- 就寝前に何も食べなかった (場合を考えてみて、、、)

## 1週間後

- 血液が(同じく)さらさらになった
- 食べなくてもさらさらになったのだから当然



はさらさらと関係ないことになる



ここで科学的に考慮すべき大事なことは、実際に行っていた行動とは反対に、何も食べなかった場合の結果がどうであったかということです。

もしもミニトマトを食べなかった場合にも食べたときと同様に血液がさらさらになっていたのであれば、ミニトマトを食べても食べなくてもさらさらとなるのですから、効果がないと判断するのが正しいといえます。

実際には食べてしまっていたので、ここで思考してみた何も食べなかった場合のことを反事実と呼びます。

効果の有無を判断する上では、反事実を考えることが重要となります。

## 2つの状況を組み合わせないと分からない

	 を食べないと	
 を食べると	さらさら	どろどろ
	さらさら	なし
	どろどろ	逆の効果あり
		なし



つまり、1つの事実だけでは効果の有無は判断できないということです。

ここで研究者が主張したいのはミニトマトに効果があることでしたが、効果があることを正しく判断するためには反事実も考え、ミニトマトを食べた場合と食べない場合で結果がそれぞれどうであったかを観察する必要があります。

「効果あり」と正しくいえるのは、「ミニトマトを食べると血液がさらさらになったが、ミニトマトを食べなかったらならなかった(どろどろであった)場合に限られます。

# つまり、1つの事実だけでは判断できない

## 実験開始日

- ・ 激戦続きでお疲れの「中田くん」
- ・ 就寝前に  を食べた

## 三「た」論法

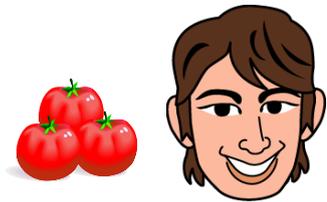
くすりを、

- ・ 飲んだら
- ・ 治った
- ・ だから効いた

## 1週間後

- ・ 血液がさらさらになった
- ・  はさらさら効果があった

はだめ！



最初にお見せした判断方法は「三た論法」と呼ばれ、科学的に正しくない判断方法です。この方法の問題は、反事実を考慮していないことにありました。

# 中田くん個人に関して

## 因果推論

-  と **さらさら** の因果関係を調べたい  
(原因と結果の関係)

- 実験開始日に **理想的なコントロール**

 を食べた中田くん

 を食べなかった中田くん

という食べたか／食べなかったかだけ異なる  が必要

- この両方の状況の結果を比較しないと、  
効果の有り無し(=因果関係)は分からない



ここで興味があるのは、ミニトマトと「さらさら」の関係です。  
すなわち、ミニトマトを原因として結果である「さらさら」がどうなるかです。

原因と結果の関係を「因果関係」と呼びます。  
この「因果関係」を評価することを「因果推論」と呼びます。

## でも

- 現実の世界では、**どちらか一方は絶対に調べられない**
- どうしたら良いか

□個人に興味があるが、調べられないので、  
**仕方なくグループについて**



のさらさら効果を調べられるか考えてみよう

- 特徴の似たコントロールが必要

□中田くんに似ている人として、



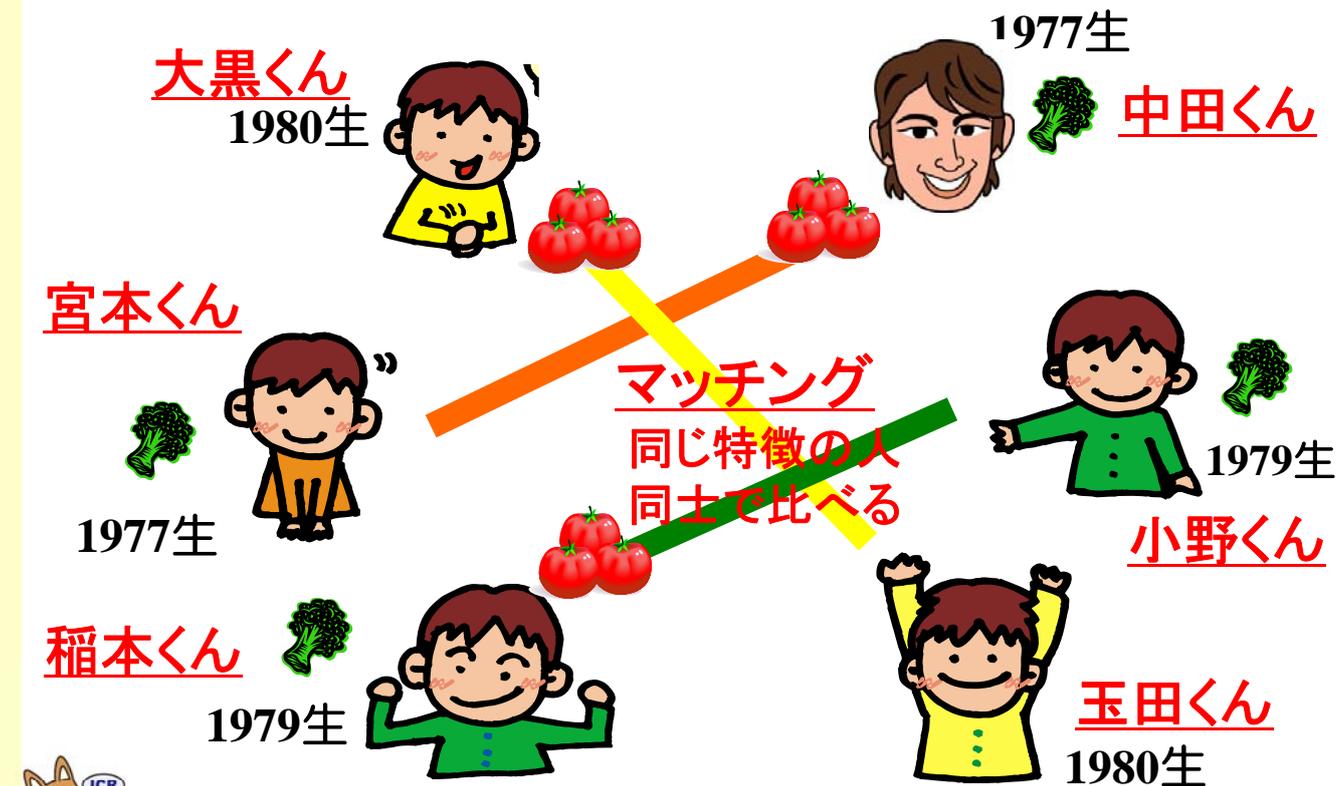
の仕事仲間と比べる



しかしながら現実の世界においては、反事実はいくまでもの想像上のものであり、絶対に観察できません。

実験当日に中田くんができることは、ミニトマトを食べるか食べないかのどちらかであり、それぞれの場合両方の結果を比べることはできません。つまり、特定の個人に対して因果推論することはできません。そこで、グループ同士を比較することを考えます。

# ブロッコリと年齢が等しい同士で比べる



グループを比較するためには、交絡が生じないように、似た特徴をもつ者同士を比較する方が適切であるといえます。

例えば、中田くんは1977生まれでブロッコリを食べていましたので、これらに関して同じ特徴を持つ宮本くんと比べる方が他の誰と比べるよりも適切でしょう。ブロッコリを食べていない人と比べたら、先ほどと同様にブロッコリによって生じた交絡によって原因が歪められてしまうことが考えられるためです。

ここでは、他のメンバーに関しても年齢とブロッコリの摂取状況が合う者同士を二人一組にして、片方にミニトマトを摂取してもらい、もう一方には摂取してもらわないことにします。

このような研究方法をマッチングといいます。

この場合、年齢やブロッコリが交絡因子であったとしても、同じ特徴の者同士で比べることで調整され、交絡が生じません。

## でも、ブロッコリと年齢だけでは

- ブロッコリと年齢だけで、**ヒトの特徴はとても表せない**  
他にも交絡因子の候補として 単純で無い

- 体格(身長、体重、)
- 運動能力
- 居住地域
- 血圧
- もっと前に食べたブロッコリ

、、、 特徴がたくさんありすぎて、  
本当は全部マッチしたいができない



しかしながら、年齢とブロッコリだけが交絡因子であるとは限りません。交絡因子の候補は他にも複数挙げられます。

また、交絡因子の候補を具体的に挙げていくと、幾つもありすぎて全てをマッチングによって対応することが不可能であることが分かります。

## そこで、

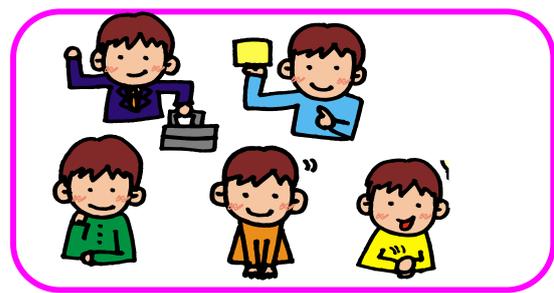
- ランダム化 randomization

- ランダム化によってグループ分け(割り付け)を決める

- ミニトマトを食べるかコントロールとなるかを  
(ここではブロッコリや年齢は全く考慮せずに)  
確率に基づいてランダムに決める



ミニトマト・グループ



コントロール・グループ



そこで互いに似たグループをつくる唯一の方法として、「ランダム化」と呼ばれるものがあります。ランダム化とは、何らかの確率に基づいてグループを割り付ける方法です。

ここでは、各被験者がミニトマトを摂取する「ミニトマト・グループ」と摂取しない「コントロール・グループ」のいずれになるかを、確率に基づいて決めます。この際被験者は、例えば2分の1の確率でいずれかのグループに割り付けられるとした場合、ほぼ均等な人数のグループが2つできることになります。

# ランダム化によって

- 食べない場合の結果(さらさら割合)が平均的に等しくなるグループをつくれる
  - つまり、個々の特徴に関しても、平均的に似た特徴をもつグループ同士となることを期待できる



ミニトマト・グループ



コントロール・グループ

平均年齢も平均身長も平均体重もブロッコリ摂取割合も、、、  
何でもグループ間で、人数が増えるにつれて似てきそう



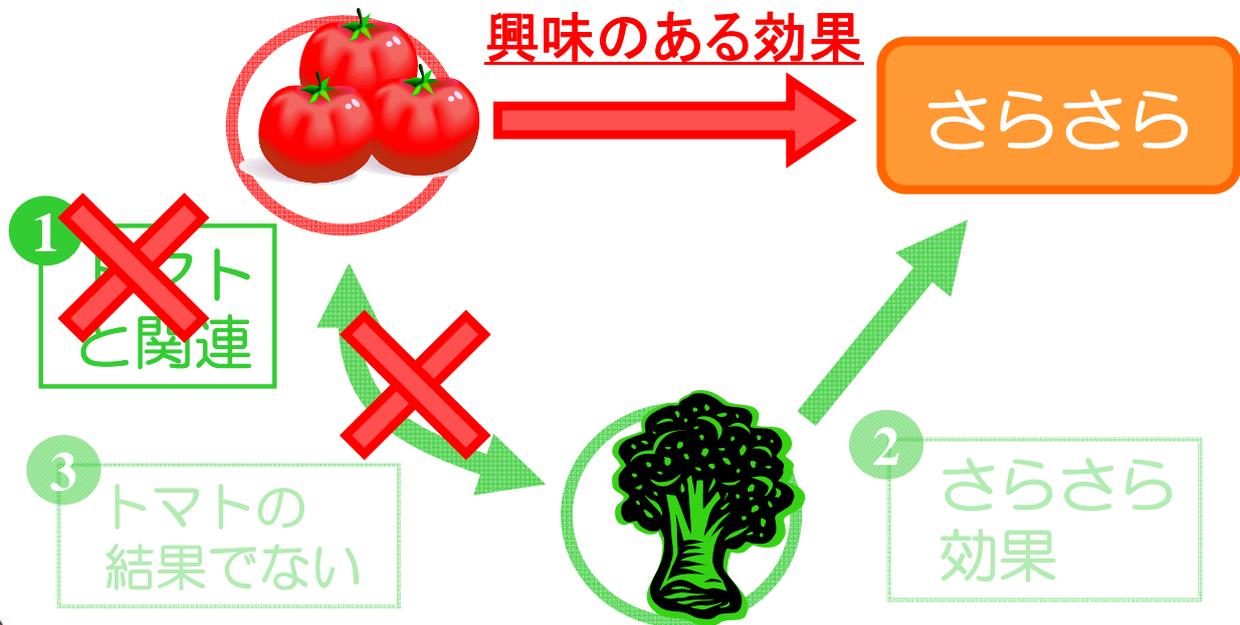
この、ランダム化を行うことにより、両グループともミニトマト食べないとした場合の結果が等しくなることが期待されます。

例え無数に交絡因子があったとしても、確率のみに基づいてグループを決めれば、平均的に似た特徴を持つグループ同士になることが期待できます。

# ランダム化によって交絡を断ち切る



と交絡因子の関連がなくなる(平均的に似る)

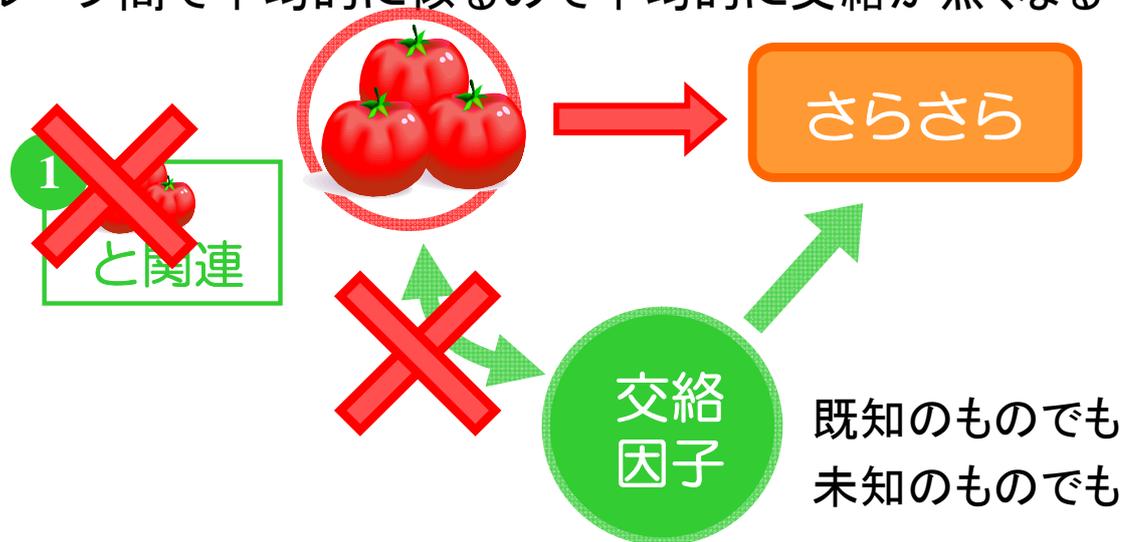


ランダム化を行うことにより、先の交絡の3条件のうち①について、期待的に満たさないようにすることができます。

ブロッコリを例にすると、ブロッコリを食べている人でも食べていない人でも、それぞれ2分の1の確率でグループを決めることにより、ブロッコリとミニトマトの関連を断ち切ることができます。

# ランダム だから既知でない交絡も断ち切る

- 実験前には交絡因子かどうか分からなかったもの、納豆でも、黒豆でも、青汁でも、それ以外でも、グループ間で平均的に似るので平均的に交絡が無くなる



また、ランダム化は、既知でない交絡因子との関連も断ち切ることができます。

他の交絡因子として納豆、黒豆、青汁などの摂取割合はここまで全く考慮していませんが、ランダム化を行うことによって、それら考慮していなかった因子との関連も断ち切ることができます。

例えば、胃がん罹患に影響を与えるピロリ菌の存在自体もつい最近まで分かっていませんでした。

ランダム化は、実験前には分からず、将来、例えば20年後、30年後に初めて分かる交絡因子に関しても①の条件を断ち切ることにより関連を断ち切ることができ、非常に大きな利点を持ちます。

## 因果推論をする際の交絡への対抗手段

- 可能であれば、最良の手段であるランダム化
  - 交絡が明らかな場合、マッチングしてからランダム化すること(層別ランダム化)もある
- 不可能であれば、**交絡因子となりうるもの全て**を測定しかつ適切に調整する
  - 層別解析(サブグループ解析)、モデル(ロジスティック回帰、Cox回帰、O $\Delta$ 回帰、...)による調整
    - 交絡因子全てを測定できているか
    - 本当に適切に調整できているかは神のみぞ知るところ
      - 非ランダム化研究の統計解析は生物統計家でも難しい解析
  - ランダム化研究よりエビデンスとしての質は**劣る**



以上のように、因果推論を行う上では、ランダム化は非常に有用な方法です。

そして、ランダム化が不可能な場合には、交絡因子となりうるもの全てを測定して調整する、という方法があります。しかしながら、交絡因子を全て測定できているか、適切に調整できているかどうかは神のみぞ知るところであり、この方法はランダム化した場合よりもエビデンスとしての質は劣るといえます。



## スタディのプロトコールを書こう！

### 因果推論とランダム化を踏まえて

- 帰無仮説「 にさらさら効果がない」
- 対立仮説「 にさらさら効果がある」
  - 臨床的に意味のある最小の差(臨床的最小有意差)は30%
- 視聴者100人をランダム化によって
  - 毎日眠る前にミニトマト3つ摂取するグループ(50人)
  - ミニトマトを摂取しないコントロールグループ(50人)のどちらかに割り付ける
- 一週間後、血液がさらさらかどうかを調べる
- グループ間でさらさら割合を比較する
  - $\alpha=5\%$ で検定、30%の差に対する検出力約85%



ここまでで説明した内容を踏まえ、実験方法に関して再検討を行い、プロトコールを書いてみましょう。



# 健康関連TVショーの正しい視聴法

## • ICR-Biostatistics版チェックリスト

- 実験の仮説(帰無仮説・対立仮説)は明確か
- $\alpha$  エラーを考慮しているか
  - P値を使って仮説検定しているか
- 検出力(または $\beta$  エラー)を考慮して  
必要なサンプルサイズを事前に定めているか
- ランダム化がなされているか
  - 非ランダム化の場合、全ての交絡因子を測定し  
適切に考慮されているか
- 事前にプロトコールで実験方法を全て規定しているか



最後に、本講義の内容を復習するために、「健康関連TVショーの正しい視聴法」を例としてチェックリストを示します。

皆さんが関わっている研究についても、チェックしてみてください。