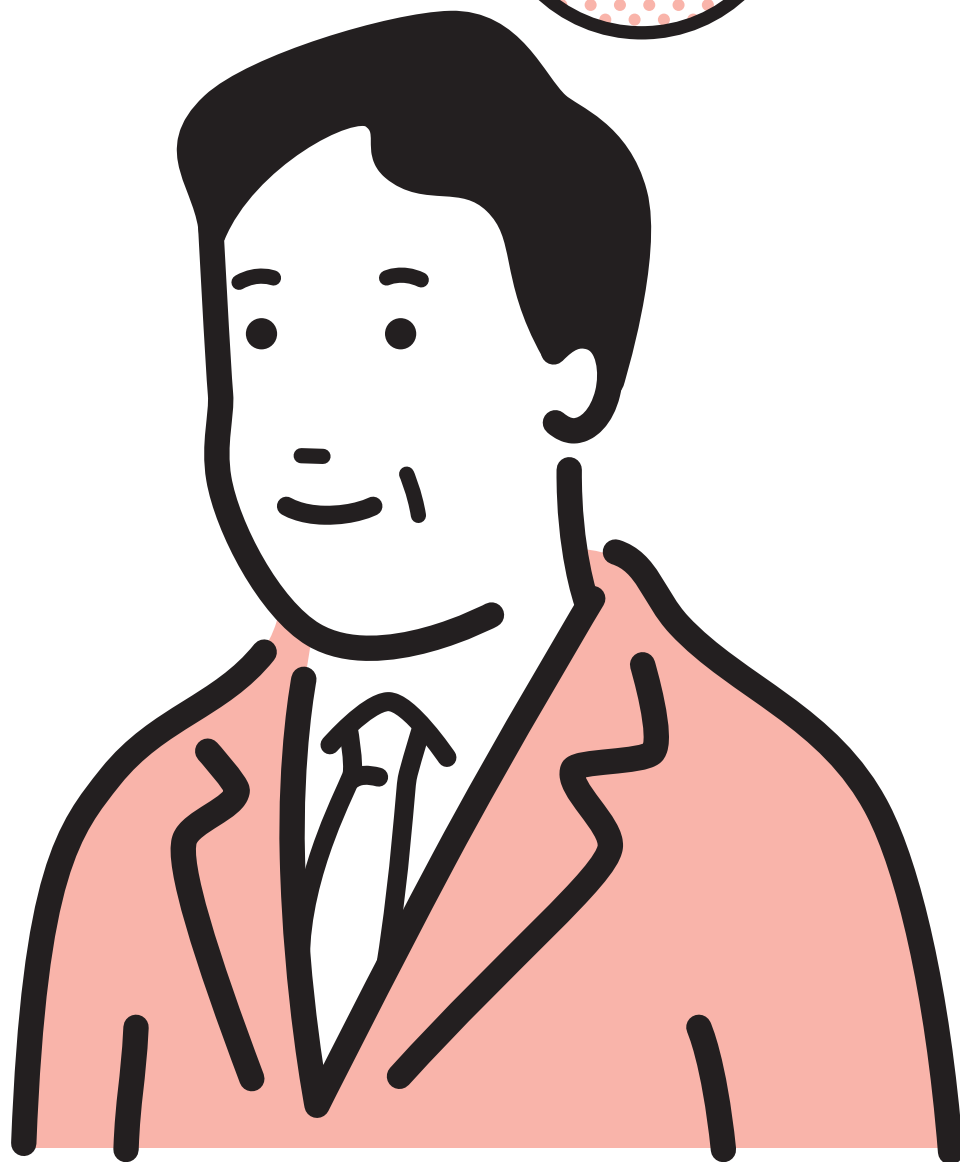


エビデンスを活かした健康づくり編



vol.3

1. データに基づいて健康づくりを企画する意義

- データから因果関係を考えるということの基本
- エビデンスレベルとリサーチデザイン

40

2. 先行研究を読み解く注意点

- 個人単位でのランダム化比較実験
- 集団単位でのクラスターランダム化比較実験
- ランダム化を行わない準実験研究・自然実験
- 調査による観察研究
- 統計的仮説検定の考え方

49

3. 健康づくりに関する倫理的配慮

57

本テキストブックについて

エビデンスに基づいて健康づくりを考えましょう、というのは言い換えるなら因果関係をよく理解した上で的確に働きかけていきましょう、ということになります。ここでは、そのために必要な「データを集めて取り組みの効果を評価する」というプロセスについて解説します。

※新型コロナウイルス感染症の感染対策を行った上で施策を実施してください。



データから因果関係を考えるということの基本 データで効果が検証された 取り組みを参考にしましょう

因果関係を理解して、的確な働きかけを！

「エビデンス※に基づいて健康づくりを考える」というのは、言い換えるなら
「因果関係をよく理解した上で先行研究を読み解く」ということになります。

※ 医学や公衆衛生の分野では、エビデンスとは科学的根拠のことを指します。詳しくは、vol.2 エビデンスに基づく健康づくり編 p22をご参照ください。

例

盛況な健康教室の事例

参加した住民は満足そうで、帰り際に「健康教室のおかげで健康になれそうだ」と好意的なコメントをくれた

Q. この健康教室は果たして参考にした方がよいのでしょうか？

A. 残念ながらこれだけでは全くわかりません。
具体的に健康にどのような効果があったのか、
適切なデータを見る必要があります。



まず「参加者が健康になれるかどうか」という結果について、この時点では不確かなことしかいえません。本人は何となく「健康になれそう」という感覚があったとしても、ただの思い過ごしかもしれません。

「ただの思い過ごし」であっても本人が良ければそれでよし、という考え方もあるかもしれませんが、こうした健康に関する錯覚は「良かれと思って逆効果」という有害な事態を引き起こす可能性があります。例えば極端な話、「自分は健康になったのだから少しぐらい暴飲暴食をしても大丈夫だろう」とか「自分は健康になったのだから少しぐらい運動不足になっても大丈夫だろう」と参加者が錯覚してしまうのであれば、健康教室が実際には参加者の健康にとって有害だったという危険性すら考えられます。

したがって、病気と関係しうる身体的指標が改善したり、健康に関係しうる行動が変化したりするのかどうかと、いった、具体的な健康指標についてはデータをもとに検証する必要があります。

病気と関係しうる身体的指標

● 血圧 ● 血糖値 など

健康に関係しうる行動

● 喫煙 ● 身体活動 ● 食生活 など

データから因果関係を考えるということの基本

データで効果が検証された取り組みを参考にしましょう



「因」と「果」を把握することが必要

前述の短期的に変化する指標は因果関係における「果」、つまり結果側の話です。最低限、結果については把握する必要がありますが、健康教室に参加した人の健康状態（結果）が改善していたとして、その原因が果たしてこの健康教室だといえるのかという「因」についても、可能な限り把握できるようにする必要があります。

例

健康教室に参加した人の
その後の健康状態についての
データを報告

地域の平均的な住民と比べると、
参加者は全般的に健康状態が良かった

Q. 健康教室は良い効果をもたらしたでしょうか？

A. このデータだけでは良い影響をもたらしたのかを
判断することはできません。「データの質」が重要です。

一見すると、健康教室は良い効果をもたらしたように思えるかもしれませんが、この結果だけでは、例えばもともと健康状態の良い住民が、その健康意識の高さ故に健康教室に参加しただけ、という可能性も考えられます。つまり、健康状態の良い住民は健康教室に参加してもしなくても全く変わらず健康状態が良く、健康状態の悪い住民はそもそも健康教室に参加しておらず、仮に参加したとしても特に健康状態が良くなったりもしない、という状況も考えられます。仮にそうだとすると「健康教室に参加した住民はその後、地域全体の平均と比べると健康状態が良い」というデータは得られてしまうのです。

もし何か健康教室などの取り組みが「原因」となって、健康状態という「結果」が改善するならば素晴らしいことです。しかし、ただデータがあれば良いのではなく「質の高いデータ」がなければ、そうした取り組みが健康状態に良い影響をもたらしたのか、全くの無駄だったのか、ともすれば逆に有害だったのかを判断することはできません。可能な限り、本当に健康に良い影響のある取り組みを明らかにし、そうした効果を実証した成果を社会全体で共有し、効果の確実な取り組みを進めていくために「質の高いデータを集めて取り組みの効果を評価する」というプロセス、すなわちサーベイデザイン（どのようにデータを集めたか）がとても重要になります。

POINT

健康づくり事業を考える際には、そのやり方が「質の高いデータを集めて
取り組みの効果が検証されている」ものかどうかによく注意しましょう。

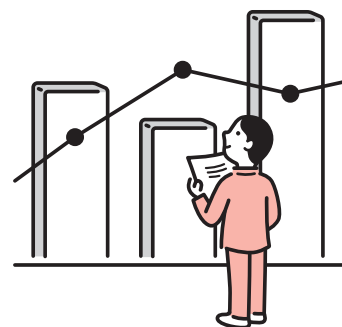


エビデンスレベルとリサーチデザイン

データの質を確認するポイント

① 十分な人数のデータはあるか

実際に効果のある健康づくり事業かを判断するには、データの件数は重要なポイントになります。



Q. 何人ぐらいのデータがあればエビデンスとして参考になりますか？

A. 数十人から数百人ぐらいがひとつの目安になります。

例えばひとりから数人に対して「ある取り組みに参加した結果、健康状態がどのように変化したか」というデータを正確に集めて報告するのもエビデンスとして参考にならないわけではありません。こうしたデータの取り方、報告の仕方はケーススタディなどと呼ばれ、学術論文に研究成果が発表されることがあります。

しかし、少人数のデータだけでは判断しにくい部分があります。ケーススタディの限界として「この少人数の参加者がたまたま特別に健康状態が改善しやすい人だったのではないか？」や「たまたま取り組みの参加前後に健康状態が改善しただけで、参加しなくても健康状態が改善した可能性はないのか？」などについてデータから判断できないことが挙げられます。

こうした「たまたま～」という限界に 대응するためには統計的な分析が必要になりますが、その際に必要なデータの目安として数十人から数百人程度の情報があれば、これが「たまたま」なのかどうかを判断しやすくなります。



「それだけでは判断しにくい部分」のことを「限界」といいます。

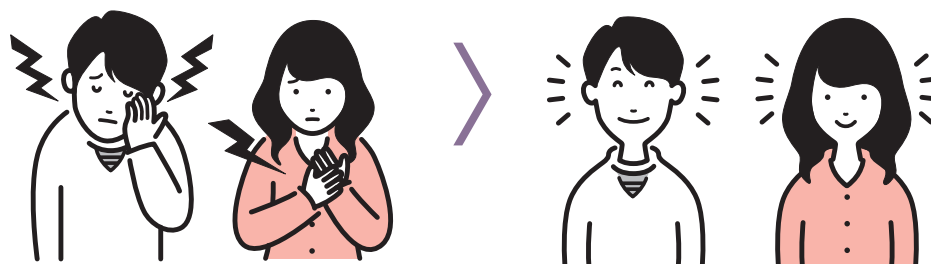
エビデンスレベルとリサーチデザイン

データの質を確認するポイント

例

たまたま一時的に体調の悪い人を数人連れてきた

安静にしていれば、数日後にはその多くの人は健康状態が回復



だからといって、体調の悪い人のほとんどは「安静にする」ぐらいで健康状態が良くなることはない。

上の例のような「たまたまじゃないか」という限界に応えるためには、統計的に十分な人数の参加者のデータが必要です。「何もなくても健康状態が改善した可能性はないのか?」という限界に応えるためには、何かしらの形で比較対照、つまり「何もしていなかった人」や「従来の健康づくり事業に参加した人」といった対照者と比べて、「効果を検証したい健康づくり事業に参加した人」の健康状態が、実際にたまたまといえるような水準を超えて良くなったのかどうかを検証する必要があります。

このような理由から、最低数十人から数百人、多い時で数万人を対象に調査を行う方がケーススタディよりも質の高いデータになります。この調査対象者の中に効果を検証したい健康づくり事業に触れた人(介入群)と、そうでない人(対照群)が含まれているのであれば、「健康づくり事業に触れた人(介入群)は統計的にたまたまといえる水準を超えて比較対照と比べて健康状態が良いといえるのかどうか」を考えることができます。



医学研究の世界では、治療や予防の取り組みなどを行った集団のことを「介入群」、それと比べる相手となる集団のことを「対照群」と呼びます。

エビデンスレベルとリサーチデザイン データの質を確認するポイント

②「選択バイアス」はあるか

データの件数以外にも、データの質を考える上で重要なポイントがあります。

**Q. 件数以外に気をつけるべきデータの質には
どういうものがあるのでしょうか？**

**A. 1度きりの調査でデータを集めるよりも、同じ人から2回以上
「ある健康づくり事業を行った時期の前と後」にデータを集めた方が
データの質は高まります。**

このような点が問題になるのは、例えば専門用語で「選択バイアス」と呼ばれる現象が存在している場合です。
下の例のような状況が想定される限り、単に1時点の調査から「健康づくり事業に参加した人の健康状態が良い」と
いった結果が示されたとしても、この健康づくり事業が原因で健康状態という結果が改善したとは限りません。
逆に「もともと健康である」という原因によって、「健康づくり事業に参加する」という結果が左右されたという因果
関係が存在しているだけかもしれないのです。

しかし、少なくとも「健康づくり事業を実施する前と後」で同じ人に健康状態を聞くことができれば、「実施前は
平均的にはほとんど健康状態に差がなかったのに、実施後は参加者の健康状態が明らかに良かった」や「実施前
にも確かにある程度平均的な健康状態に差があったものの、その差以上に実施後の参加者の健康状態が明らかに
良くなっていった」といったことがわかります。

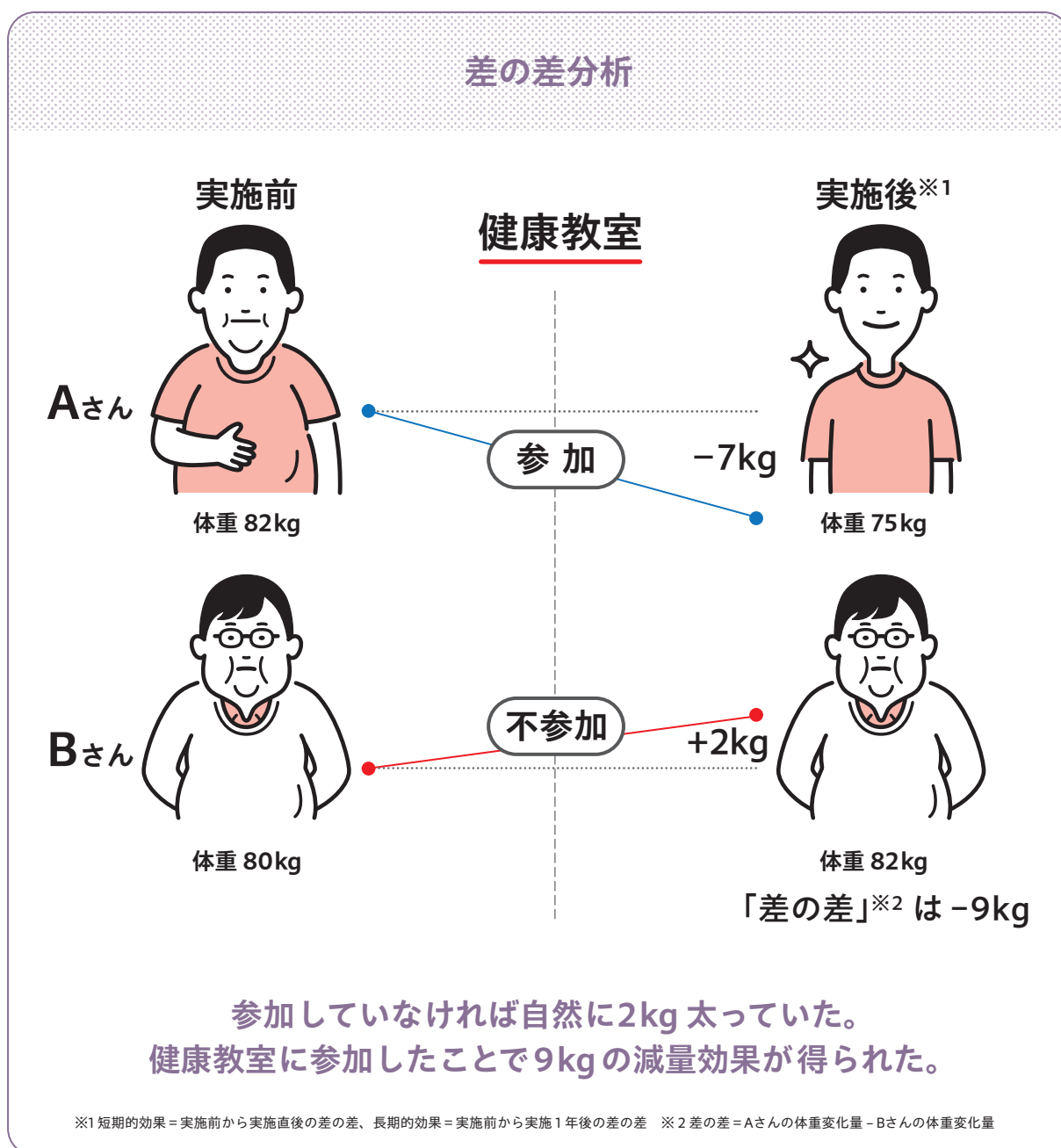


エビデンスレベルとリサーチデザイン

データの質を確認するポイント

こうした分析のやり方を「差の差分析」といいます。つまり、「健康づくり事業の実施前と実施後の差」に関する、「健康づくり事業に参加した人とそうでない人の差」を分析するため、「差の差分析」というわけです。

同一対象者に対して事業実施前後の2時点、あるいは「実施前、実施直後、実施1年後」といったように3時点でデータを収集して、短期的な(事業実施前後の)「差の差」と、長期的な(実施前から1年後までの)「差の差」を両方分析して事業評価をすることもできます。



エビデンスレベルとリサーチデザイン データの質を確認するポイント

③「思い出しバイアス」はあるか

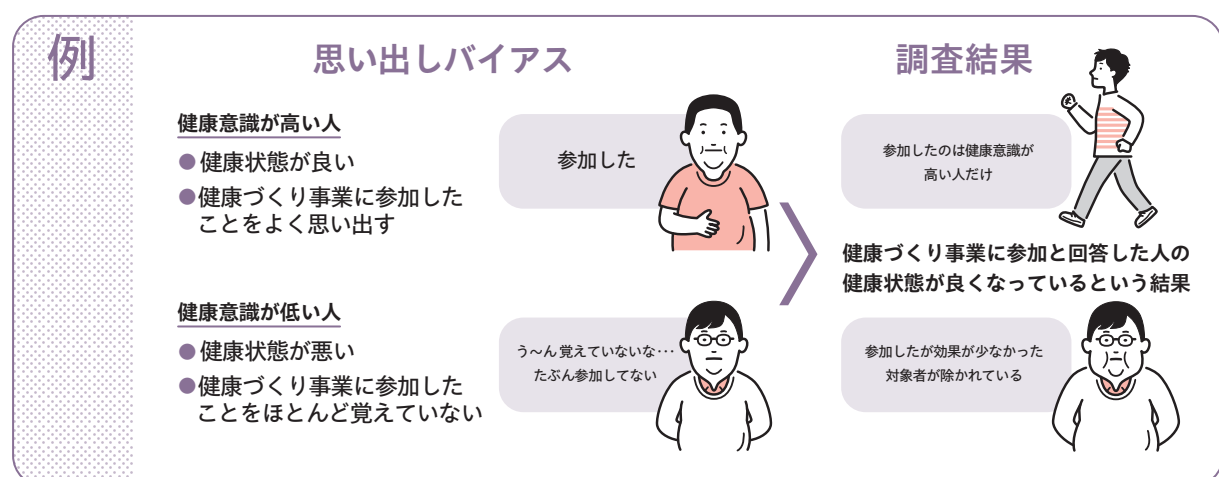
選択バイアスの他にもうひとつ、思い出しバイアスについてもデータの質を高める方法があります。

Q. 「思い出しバイアス」とはどのようなものですか？

A. 調査に回答している人が、実際には健康づくり事業に参加していながらそのことを覚えていなかったり、参加しなかったのに「何かそういうものがあった気がする」と錯覚して「参加した」と調査に回答してしまうと、調査結果をゆがめてしまう原因になります。

参加者の回答だけから「ある健康づくり事業に参加したかどうか」を判断するよりも、実際に参加したのかどうか、という事実をもとに判断した方がより正確なデータであるという考え方もできます。

ただ健康づくり事業を全体的に実施し、それに触れたかどうかを結果とともに調査する、というのではなく、一部の人に対しては確実に健康づくり事業に参加してもらい、確実にその事業に参加していなかった人と比べてどれだけ「差の差」が改善したのかを把握する方が、より質の高いデータであるといえるでしょう。このような評価方法は「実際に一部の人で行って試す」という意味から「実験研究」と呼ばれます。一方で、1時点であれ複数時点であれ、明確に「参加する介入群と参加しない対照群」に誰が該当するかを設定せずに調査を行うことを「観察研究」と呼びます。



エビデンスレベルとリサーチデザイン データの質を確認するポイント

④ランダム化は行われているか

実験研究のやり方に応じてデータの質が異なります。

Q. 実験研究にはどのようなやり方があるのでしょうか？

A. ランダム化をする研究としない研究があります。

ランダム化を行うとは、例えば「100人の研究協力者のうちランダムな半分には新しい健康づくり事業に参加してもらい、残りの半分には従来どおりの健康づくり事業に参加してもらおう」というものを指します（なお、新しい健康づくり事業に参加する介入群と、残りの半分の対照群は必ずしも半々でなければいけない、ということはありません）。

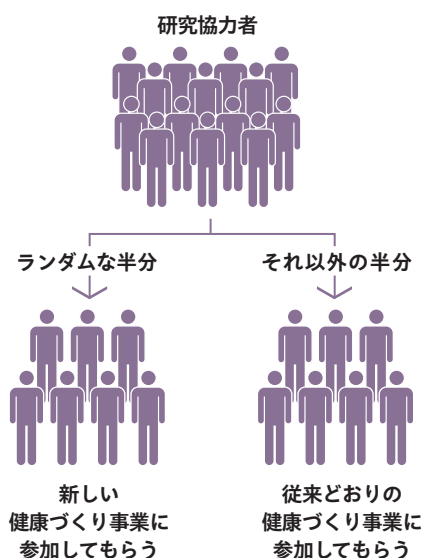
一方で、ランダム化を行わない場合の実験研究の例としては「A、Bという2つの地域に住む住民のうち地域Aの住民には新しい健康づくり事業に参加してもらい、地域Bの住民には従来どおりの健康づくり事業に参加してもらおう」という準実験研究が挙げられます。

あるいは、意図的に行ったわけではなく、「自然と実験のような状況になった」という自然実験と呼ばれるものも、ある種の準実験研究だと考えることもできます。つまり、たまたま地域A全体で健康づくり事業が行われていた時、隣接した同じような地域Bでは全くこうした取り組みが行われていなかった、という状況のデータを遡って収集・分析すれば「自然とランダム化を行わない準実験研究のような状況だった」という自然実験のエビデンスだと考えられるわけです。

ランダム化をするかしないかで、「新しい健康づくり事業に参加してもらおう人」と「従来どおりの健康づくり事業に参加してもらおう人」の差に偏り（バイアス）があるかどうかという点に大きく違いが出ます。

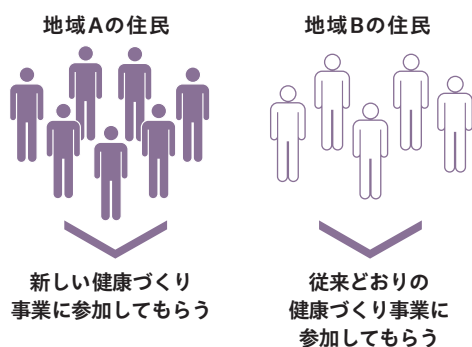
例

ランダム化比較実験



バイアスなし

ランダム化を行わない準実験研究



バイアスがないとはいえない

地域Aと地域Bの住民の間で、健康状態やその変動の仕方に大きく差があったら、新しい健康づくり事業が原因で健康状態が良くなったの？それとも単に地域Aの住民は地域Bの住民よりもこの実験期間中に健康状態が顕著に良くなっていたの？



エビデンスレベルとリサーチデザイン データの質を確認するポイント

このように、準実験研究を行うのであれば極力注意深く、比較対照（比べる相手）が「概ね同じような集団」となるよう気をつけなければいけません。

しかし、いくら気をつけても本当に「概ね同じような集団」なのかを保証することはできません。ここでランダム化を行うことができれば、例えば誰が投げてでもサイコロの出目の平均値に偏りが無いのと同じように、地域からランダムに選んだ調査対象者の回答の平均値は地域全体の全数調査の平均値と偏りがなく、またバラつくにしてもどの程度のバラつきになるかが確率的に計算可能になります。したがって、ランダムに分けられた集団の間にひとつだけ「新しい健康づくり事業に参加してもらうかどうか」という違いを作った結果、偶然のバラつきとはいえない水準で両者の健康状態に差が生じていたとしたら、「新しい健康づくり事業への参加（原因）により健康状態（結果）に差が生じていた」という因果関係がかなり正確に説明できることになります。

健康づくり事業は、個人ごとに「参加してもらうか否か」をコントロールできるものばかりではありません。

- 学校や事業所、公園といった施設単位で、いざ事業を実施しようとする
と「施設に所属していたり、利用していたりする人全体が対象になる」
といった健康づくり事業
- 施設といった形で明確な区切りがなくても、いざ事業を実施しようと
するとごく狭い特定の地域の中にいる全員が対象になるような事業

このような場合は、人ではなく施設や地域をランダムに「新しい健康づくり事業を行う介入群と比較対照とする対照群」に割り付けて、そこに所属している人やふだん利用している人の差を検証することが可能です。これをランダム化比較実験のうち特別なものとして、専門用語で「クラスターランダム化比較実験」と呼びます。



\ POINT /

データの質は様々です。「十分な人数のデータはあるか」「偏り（バイアス）はないか」「ランダム化は行われているか」を確認し、質の高いデータを参照しましょう。

個人単位でのランダム化比較実験 まずランダム化比較実験を 参照しましょう

シンプルでパワフルなランダム化

もしあなたが実施しようとする健康づくり事業が、対象者一人ひとりに対して「介入群となるか」「比較のための対照群となるか」を制御可能なものなら、まずは参考になりそうな過去のランダム化比較実験の成果を探してみましょう。

※ランダム化比較実験については、Vol.2 エビデンスに基づく健康づくり編 p25もご参照ください。

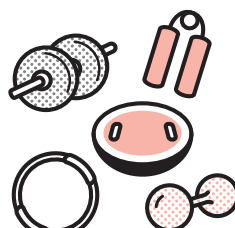
対象者一人ひとりに対して 介入群となるか対照群となるかを制御可能な取り組み



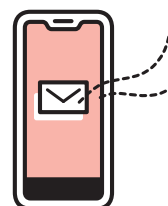
直接会話 / 電話 /
オンライン面談



手紙や紙の
資料を郵送



健康に役立つ器具の
貸し出し / 譲渡



メール / アプリ

Q. なぜ最初にランダム化比較実験を参照するのですか？

**A. ランダム化を行うことができれば、かなり正確に
「健康づくり事業（原因）により人の健康状態（結果）が良くなった」
という因果関係を示すことができます。**

評価の対象者数としては、それぞれのグループについて、最低数十人、可能なら数百人ずついる状態で、グループ間の健康状態の差が統計的に「偶然のバラつきでも得られる範囲」と考えにくいような水準であれば、おそらく「健康づくり事業（原因）により人の健康状態（結果）が良くなった」という因果関係があったのだろうと示すことになります。

個人単位でのランダム化比較実験

まずランダム化比較実験を参照しましょう

ランダム化比較実験の注意点

シンプルでパワフルなランダム化比較実験ですが、気をつけることが2点あります。

Q. ランダム化比較実験のどこに注意したらいいですか？

A. 効果のバラつきを考えて複数のランダム化比較実験の結果を参照するとともに、自分たちが実施を想定する状況や対象集団についても当てはまりそうなのかをよく注意するという点です。

ランダム化比較実験によって示される効果は常に一定ではなく、ある程度のバラつきが存在しています。そうしたバラつきを超えて、「おしなべてポジティブな効果があるといえそうなのか」という点をみましょうというのが、ひとつ目の注意点です。システマティック(系統的)レビューと呼ばれる形式の論文や報告書には、こうした複数のランダム化比較実験をまとめ、全体的な効果の程度について言及しているものがあるので、できれば自分の行おうとするタイプの健康づくり事業について、そうしたまとめがないか探してみましょう。

また、ランダム化比較実験の効果には単純なバラつき以外に、「特定の状況では効果が高い」「特定の対象者では効果が低い」といった効果の異質性が存在しているかもしれません。そうした点もしばしばシステマティックレビューの中で言及されていますので、よく注意して読んでみましょう。そこから「どんな状況・対象者でもある程度効果が見込めそう」なのか、それとも「自分の想定する状況・対象集団の効果は見込めなさそう」なのか、「現時点であまりエビデンスが揃っていない」のかを検討することも重要です。

\ POINT /

バラつきと異質性に注意し、参考になりそうなランダム化比較実験の成果を探し、健康づくり事業の企画に役立てていきましょう。

集団単位でのクラスターランダム化比較実験 集団単位での制御しかできない 事業はクラスターランダム化 比較実験を参照しましょう

クラスターランダム化比較実験

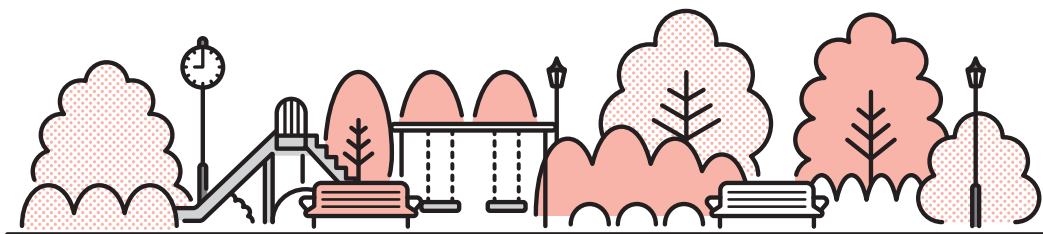
もしあなたが実施しようとする健康づくり事業が、例えば学校や事業所、公園などの施設単位あるいは小さな地域単位で「やるなら施設や地域に所属していたり、利用していたりする人全体が対象になる」ものとしたら、複数の施設(地域)をランダムに「事業を実施する施設(地域)」と「事業を実施しない施設(地域)」に割り付けるクラスターランダム化比較実験についてのエビデンスがないか探してみましょう。

Q. どのような事業がクラスターランダム化比較実験に該当しますか？

A. 例えば公園を運動しやすく整備したり、職域の事業所全体で何かの取り組みを行ったり、学校の中で特別な教育プログラムを実施したり、という事業が挙げられます。

このような事業では、その中にいる一人ひとりに対して介入群となるか対照群となるかを制御することができません。強いていうなら学校については、学校単位だけでなくクラス単位で「新しい教育プログラムを受けてもらうかどうか」を制御することができるかもしれませんが、一人ひとりを制御することはしばしば困難です。

エビデンスレベルとリサーチデザインの項でも触れましたが、このように介入群となるか対照群となるかが個人単位ではなく施設や地域、集団といった単位でしか制御できないにしてもランダム化を行うことはでき、それによってかなり正確に「健康づくり事業(原因)により人の健康状態(結果)が良くなった」という因果関係を示すことができます。



\ POINT /

施設や地域単位のランダム化では、個人単位のランダム化比較実験よりも事業を実施するグループとそうでないグループの偏りが生じやすいため、ランダム化されたグループ間で、研究参加者の性別や年代、職業や居住地域などの条件に偏りがないかによく注意して、論文や報告書をチェックするようにしましょう。

ランダム化を行わない準実験研究・自然実験

ランダム化がしにくいテーマは準実験研究や自然実験を参考にしましょう

準実験研究ができないか検討する

もしあなたが実施しようとする健康づくり事業が、対象者一人ひとりに対しても、施設や地域ごとといった形でもランダムに介入群となるか対照群となるかといった状態を制御することが難しいようであれば、準実験研究や自然実験といわれるエビデンスを探してみましょう。

Q. どのような事業が準実験研究や自然実験に該当しますか？

A. マスメディアを使って多くの人にメッセージを届けるような取り組みや、自治体内の制度を大きく変更する場合は当てはまります。

健康づくり事業を行うなら、地域全体で同時に行わなければいけない、といった状況の時、この地域の中では全てが「事業を行うところ」になるため、意図的に「事業を行っていないところ」を作り出すことはできません。しかし、別の見方をすればその近くにある別の地域や、そこにある同様の施設は「事業を行っていないところ」という比較対照と見なせるかもしれません。このような状況のデータを収集し、適切に分析した結果は、ランダム化こそしていないものの、ある種の実験のような状況という意味で準実験研究と呼ばれます。あるいは、たまたま自然とそのような状況だったということで自然実験と呼ばれることもあります。

準実験研究や自然実験についても、複数のエビデンスを参照し「全体として効果がありそうか」「自分が実施を想定する状況や対象集団にも効果があるといえそうか」を考えることが重要であるのは、ランダム化比較実験と同様です。

また、ランダム化をしていない分、クラスターランダム化よりさらに注意深く「事業に触れた介入群とそうでない対照群の間に偏りがいないか」という点をチェックする重要性が増します。特に、効果が見込まれるであろう健康状態について事業実施前に大きな偏りがある時は、単に「健康状態の良い人が多い地域で事業を実施した」という可能性について、よく確認しましょう。「実施前と実施後の差」に関する「実施した地域としていない地域の差」を見ると「差の差分析」を行っている研究や報告書も多数ありますが、もしこうした分析をしていなかった場合、本当に事業が原因で健康状態の結果が改善したという因果関係が考えられるか、よく注意してエビデンスを吟味しましょう。

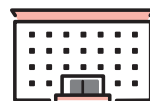
● 所属する自治体

私の所属する自治体では
遊歩道を整備



● 隣の自治体

遊歩道は
整備していない



隣の自治体は
「事業を行っていない
ところ」になるのね。

POINT

準実験研究や自然実験は、ランダム化をしていない分、クラスターランダム化よりさらに注意深く「事業に触れた介入群とそうでない対照群の間に偏りがいないか」という点をチェックすることが重要です。

調査による観察研究

観察研究の読み解き方

観察研究とは？

ランダム化比較実験も準実験研究や自然実験のエビデンスも見つからなければ、最低限何かしらの過去の調査データに基づいて事業の効果が見込めそうなのかどうかを確認してみましょう。

観察とは「積極的に実験を行わず、そこにいる人たちの情報を取得するだけ」という意味です。

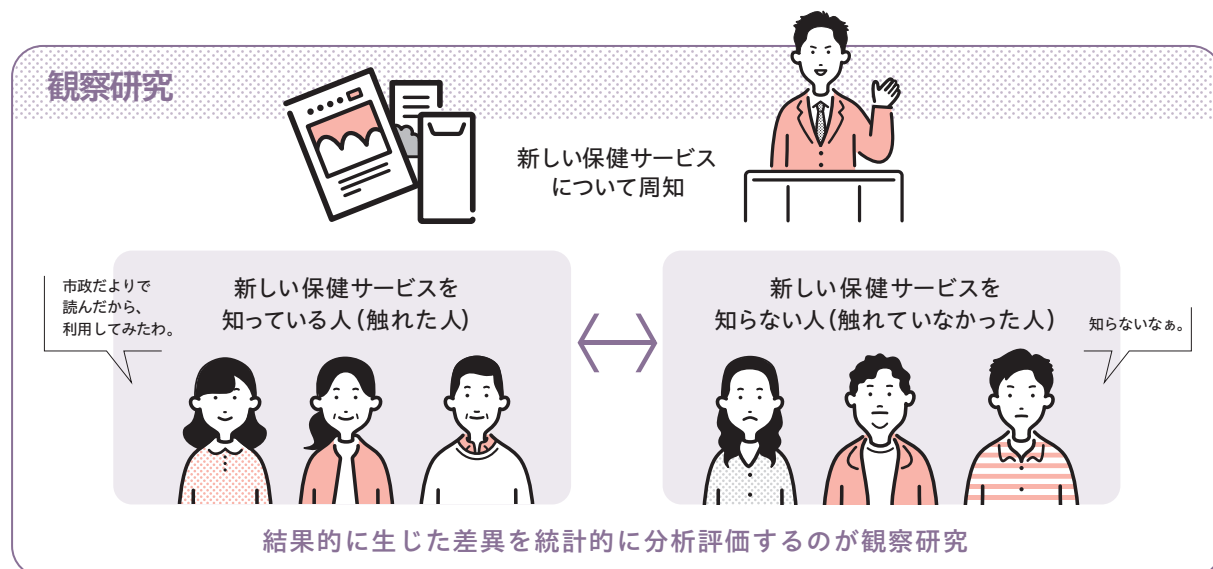


Q. 準実験研究も難しい時には何を参考にしたらいいですか？

A. 「特定の状況にあった人」と「そうでなかった人」の差を分析する観察研究を考慮しましょう。

もしあなたが実施しようとする健康づくり事業が、やるとするならば地域全体で同時に行わなければいけない、とか、自分たちの地域には健康づくり事業の対象となる施設がひとつしかない、といった性質のものであるなら、基本的には他地域を比較対照とした準実験研究や自然実験を探すのがおすすめです。しかし、どうしてもランダム化比較実験どころか準実験研究や自然実験のエビデンスも見つからないのであれば、最低限何かしらの過去の調査データに基づいて事業の効果が見込めそうなのかどうかを確認してみましょう。

このような場合、「(自分たちが実施を想定するような)健康づくり事業に触れた介入群」と「そうでない対照群」に誰が該当するかは、自分たちで予めコントロールすることはできません。例えばどれだけ熱心にマスメディアを使って何かのメッセージを呼びかけても、行政の新制度についての周知徹底を心がけても、短期間で全ての住民が「そのメッセージに触れた」とか「新制度を知って申込みをした」ということは考えにくく、どうしてもバラつきは出てきます。そこでたまたま生じた「(意図的に制御はできないが結果的に生じた)介入群」と「(同じく結果的に生じた)対照群」の差を分析するのが観察研究の考え方です。



調査による観察研究

観察研究の読み解き方

観察研究の難しさ

観察研究では、調査の結果、仮に「結果的に介入群となった人」が統計的にたまたまといえる水準を超えて「結果的に対照群となった人」より健康状態が良かったからといって、必ずしも健康づくり事業（原因）により、健康状態（結果）が改善したという因果関係が示せるわけではありません。

Q. 観察研究で気をつける点がありますか？

A. 可能な限り、同じ人に対して「健康づくり事業を実施する前と後」の健康状態と、健康状態および健康づくり事業への触れやすさに関する情報を網羅的に取得しておくことが推奨されます。

観察研究では、健康状態および健康づくり事業への触れやすさに関する情報をどれだけ可能な限り取得できるかの重要性が、クラスターランダム化比較実験や準実験研究よりもさらに大きなものになります。ランダム化していれば、健康づくり事業に触れた介入群とそうでない対照群の違いはそれほど気にしなくてもよくなりますが、観察研究では研究者の想定しきれないところで両者の違いが生じているかもしれません。例えば観察研究における「結果的に介入群となった人」と「結果的に対照群となった人」の間で、性別、年齢、職業などの基本的な属性のほか、健康に関する意識や、行政サービスへのアクセスのしやすさ、ライフスタイルや時間的余裕、地域社会とのつながりの強さなどが大きく異なっている可能性もあります。

そのため、どのような条件が「結果的に介入群となった人」と「結果的に対照群となった人」の間で偏りうるのか、データ分析上の工夫としてそれらの偏りにどう対処されたのかは、先行研究を読み解く上でよく注意しなければなりません。



観察研究でも
「選択バイアス」や
「差の差分析」の
考え方は同じです。



POINT /

観察研究で事業の有望さを検討する場合は、ぜひ統計学や疫学に詳しい専門家をまじえて議論することをおすすめします。また、もちろんひとつの観察研究だけではなく複数の（同じような健康づくり事業に関する）観察研究を参照・比較することが重要であるということでもあります。

統計的仮説検定の考え方

ただの誤差といえるような効果かどうかを判断しましょう

統計的仮説検定とは？

健康づくり事業の評価では、事業に触れた人とそうでない人との間に「統計的にたまたまとは考えにくい水準の差」が生じるかどうか重要です。

Q. 「統計的にたまたまとは考えにくい水準の差」かどうかはどうやって判断するのですか？

A. 統計的仮説検定を行います。

ほんの数人ずつの「健康づくり事業に触れた人たち」と「そうでない人たち」の間の平均的な健康状態に差があっても、それはたまたまだろう、ということは直感的にわかります。一方、数万人ずつが含まれるグループの間で平均的な健康状態に差があれば、それは「たまたま」というのはいけなような気がします。

これらをより正確に判断するためのものが、統計的仮説検定と呼ばれるものです。統計的仮説検定の基本的な考え方としては、まず「同じ条件で繰り返しデータを集めたとき、事業に触れた人と触れていない人の平均的な健康状態に差はない」という仮定のもとで、「両者の健康状態に差がある」というデータが得られる確率を計算します。この確率はprobability (確率) という英語の頭文字をとってp値と呼ばれますが、これが大きい場合は「健康状態に差がある」のはよくあることとして、「現時点ではたまたまといえる程度の差」だと判断できます。一方で、このp値が小さいものであれば最初に立てた「同じ条件で繰り返しデータを集めたとき両者の健康状態に差はない」という仮定のもとでは、そうそう得られないような「健康状態に差がある」データが得られていることになり、これは「たまたまといえる程度の差」とは考えにくいという結果になります。なお、慣例的にはp値の目安として5%を下回っていれば「たまたまといえる程度の差とは考えにくい水準」と判断されます。この状態のことを専門用語で「統計的に有意な差」と表現します。

統計的仮説検定の考え方

ただの誤差といえるような効果かどうかを判断しましょう

効果とバラつき

データを得る研究参加者が多ければ多いほど「統計的に有意な差」は得られやすくなります。また、同じ研究参加者の数でも、もともとの値のバラつきが小さかったり、効果（つまり事業に触れた人と触れていない人の健康状態の差）の程度が大きければ大きいほど、「統計的に有意な差」と判断されやすくなります。

Q. 「統計的に有意な差」がないと、その事業の効果は「ない」のですか？

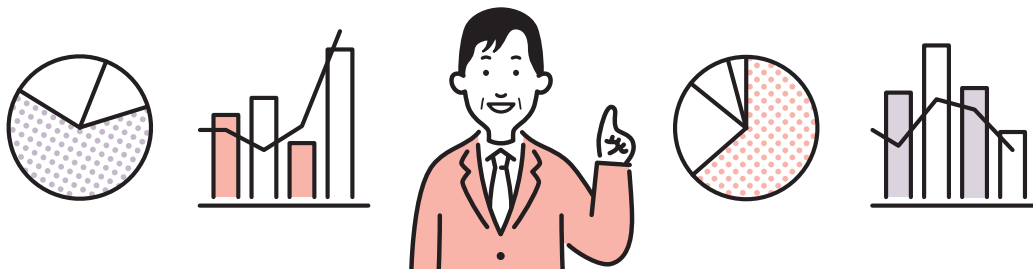
A. 「バラつきが大きすぎてその程度の人数では検証しきれない」というだけかもしれません。

データを得る人数のことを専門用語で「サンプルサイズ」といいます。



もともとのデータのバラつき方と事業の効果について例を挙げると、「もともと ± 10 程度のバラつきしか存在しない指標について、健康づくり参加者は平均すると数値が20も改善した」のであれば、比較的少人数での検証でも有意な差と判断できる可能性が高くなります。一方で、「もともと $\pm 1,000$ ぐらいのバラつきが存在している指標について、健康づくり参加者は平均すると数値が20も改善した」という結果が有意な差といえるかは、かなりの大人数で検証しなければわかりません。

つまり、0.05より大きなp値が報告されていたからといって、その事業の効果が「ない」というわけではなく、「バラつきが大きすぎてその程度の人数では検証しきれない」というだけなのかもしれません。したがって、統計的に有意な効果を示すエビデンスが見つからなかったからといって、同様の事業を行うことが必ずしも無駄とは限りませんが、他に統計的に有意な効果が見られているものがあればそちらを優先した方が良いでしょう。



POINT

実施する健康づくり事業が「統計的にたまたまとは考えにくい水準の差」なのかどうか、よく注意して確認してみてください。

健康づくりに関する倫理的配慮

生命倫理の4原則に則って考えましょう

公正な健康づくり事業を

「人々を健康にする」というのは素晴らしいことですが、その一方で影響力も大きく、良かれと思って悪影響をもたらしてしまうことや、行政が不平等に一部の住民にだけメリットを与えるというリスクもあるため、倫理面での配慮も重要なポイントになります。

Q. 倫理的な配慮はどのように考えれば良いのでしょうか？

A. 「生命倫理の4原則」および「人を対象とする生命科学・医学系研究に関する倫理指針」に基づき、非倫理的な取り組みとはならないよう慎重な研究体制で臨むことが求められます。

「生命倫理の4原則」とは、医療倫理の大家であるトム・L・ビーチャムとジェイムズ・F・チルドレスが『生命医学倫理』という古典的名著¹⁾において提唱した概念であり、医療の世界で倫理の原則としてよく用いられます。以下の4つからなります。

- 自律尊重の原則 …… 「自分の意思で決定することのできる人が、選択する自由がある状況で、自身のことを自分で決め、行動すること」
- 無危害の原則 …… 「患者に危害を及ぼさず、今ある危害や危険を取り除き、予防すること」
- 善行の原則 …… 「(医療行為を行う)相手に対して相手が考える最善をなすこと」
- 公正の原則 …… 「相手を平等かつ公平に扱うこと」

これらの基準から考えると、自律尊重の原則から、健康づくり事業において、相手の行動を強制することは基本的に許されません。無危害および善行の原則から、健康に有害な行動をかえって増やすような取り組みはしてはならず、そうならないよう最善を尽くすべきだ、といえます。また、すでに健康面でも経済面でも恵まれている立場にあるごく一部の住民がさらに健康になる、といった健康づくり事業も公正の原則から推奨されないと考えられます。

先行研究をしっかりと調べて、実施しようとする健康づくり事業が不公正なものでないか、危害のあるものでないか、と検討することは骨の折れることですが、一方でそうした下調べも含めて最善の事業を計画するというのは「善行の原則」に合致するものです。

厚生労働省では「人を対象とする生命科学・医学系研究に関する倫理指針」を定めており、健康づくり事業に関する研究を行う際にもこの指針に基づき、非倫理的な取り組みとはならないよう慎重な研究体制で臨むことが求められます。一般的な事業の範囲を超えて、自分たちで新しい取り組みを考案し、その効果を検証したいということであれば、ぜひ地域にある大学や研究機関などと共同で倫理指針にそった研究体制を構築の上、適切な倫理審査を受けることをおすすめします。

\ POINT /

効果があるのかわからないのか不確かな健康づくり事業を企画して倫理的問題を招かないようにするためにも、エビデンスに基づくというプロセスはとても重要です。

1) トム・L・ビーチャム, ジェイムズ・F・チルドレス. 生命医学倫理. 第五版. 立木教夫, 足立智孝, 監訳. 千葉: 麗澤大学出版会; 2009.