

[総 説]

研究論文の公表や抄読に必要な統計学・疫学の基礎知識
～因果推論を中心とした基本的な考え方の紹介～

比江島欣愼

東京医療保健大学大学院医療保健学研究所

(平成 27 年 5 月 7 日受付)

医療分野で行われる研究・調査において、その興味の多くは因果関係を示すことにあります。そのために、研究・調査をデザインし、データを収集・分析し、因果関係の有無を推論します(因果推論)。本稿では、まず、因果推論に必要な統計学に関する基本的な考え方を紹介します。その後、どのようにして因果関係を推論するのか、その考え方の基本となるモデル、カウンターファクチュアルモデルを紹介し、ランダム化比較研究がどうして因果推論に最適な研究デザインであるのかを説明します。最後に、因果推論の際によく用いられる検定のロジックを解説し、検定手法の選択方法を紹介します。

Key words: 統計手法, 検定, 因果推論, カウンターファクチュアルモデル, ランダム化比較研究 (試験)

1. はじめに

医療分野における研究や調査においてデータ収集が行われる場合、因果推論をその目的とすることが多いです。薬や治療と症状改善との関連、抗菌剤や感染対策と菌やウイルスの死滅との関連、生活習慣と疾病発生の関連など、多くの関連性の確認は因果関係をデータから推論する形で議論されます。本稿では、因果推論の為に必要な統計学に関する基本を紹介し、因果推論の考え方と理想的な研究デザイン、因果推論に用いられる検定のロジックとその手法選択などについて入門的な解説を行います。

2. 集団を対象にする際の基本的な考え方

データ収集をともなう研究や調査では、多くの場合、ある集団の様子に興味があります。この興味の対象となる集団をここではターゲット集団と呼ぶことにします(一般の統計学のテキストでは母集団と呼ばれ

ています)。

さて、ターゲット集団の様子を正確に把握するにはどのようにデータを採ればよいでしょう? 集団内のすべての人からデータを採れば、正確かつ詳細にその様子を把握することができます。このような形でデータ収集を行う調査を全数調査と呼びます。しかしながら、ターゲット集団のサイズが大きい(集団に属する人がとても多い)場合、その全員からデータを採ることは容易ではありません。

そこで、全数調査をあきらめて、ターゲット集団の一部の人たちからデータを採ってターゲット集団の様子を把握することを考えます。このような調査方法を標本調査と呼びます。このときのデータ収集の対象となるターゲット集団の一部の人たちを標本と呼びます。標本から得たデータを利用して、ターゲット集団全員を調べる場合に得られる結果(全数調査の結果)を推測することが、標本調査では必要になります。

では、どのように標本の対象をターゲット集団から選べばよいのでしょうか? 標本から得たデータを利用して全数調査の結果を推測したいので、なるべく標本の対象をターゲット集団から偏ることなく選び出し、ターゲット集団の様子をよく反映した標本になるようにする必要があります。この要求に応えてくれる唯一の選出(抽出)方法がランダムサンプリングになりま

著者連絡先: (〒154-8568) 東京都世田谷区世田谷 3-11-3
東京医療保健大学大学院医療保健学研究所
比江島欣愼
TEL: 03-5799-3711
FAX: 03-5799-3713
E-mail: y-hiejima@thcu.ac.jp

す。標本の対象をターゲット集団からランダムに抽出するという方法です¹⁾。

このように、ターゲット集団からある程度の数の対象をランダムサンプリングによって抽出することで、ターゲット集団の様子をよく反映した標本が作られることが期待できます。その数が多ければ多いほど、標本の様子はターゲット集団の様子を反映します。この性質は大数の法則と呼ばれています²⁾。ランダムサンプリングによって作られた標本は、ターゲット集団のミニチュア集団と言ってもよいでしょう。この標本から採られるデータを利用することで、ターゲット集団における全数調査の結果を推測することができます(図1参照)。標本調査においてランダムサンプリングはなくてはならない手続きであることをここで強調しておきます。

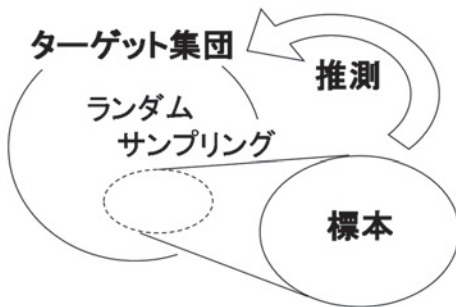


図1. 標本調査の枠組み

3. カウンターファクチュアルモデル

本節では、因果推論の基本になる考え方について解説します。因果推論とは、ある現象の発生がどのような要因と関連しているかを知るために、データを収集し、得たデータからその現象と要因の因果関係を推論することです。医療分野における興味の対象となる問題の多くが因果推論の範疇に入るといってよいでしょう。では、因果推論をどのような手続きで進めていけばよいでしょう？

3.1. 1人について考えてみる

ある人に胃ガンの発症を防ぐためにピロリ菌除菌を

行ったところ、胃ガンを発症することもなく20年を無事経過しました。この人においてピロリ菌除菌は20年間胃ガンを発症しないことの原因だった(ピロリ菌除菌は胃ガンの予防に効果があった)のでしょうか？このことを裏付けるためには、どのようなことが観察されるべきかを考えてみましょう。まず、少なくとも以下のことが観察されなければなりません。

①この人にピロリ菌除菌が行われ、20年間胃ガンが発症しない

この観察結果は因果関係を裏付けるには必要不可欠ですが十分ではありません。もし「この人にピロリ菌除菌を行わず、20年間胃ガンが発症しない」が同時に観察されたとするのでしょうか。ピロリ菌除菌をしようがしまいが胃ガンを発症しないのですから、ピロリ菌除菌以外の何かが胃ガンの発症を防いでいる(もしくは、もともと発症しない)ということになります。となると、因果関係を裏付けるためには①の観察だけでは不十分で、以下のことも観察される必要があります。

②この人にピロリ菌除菌が行われず、20年以内に胃ガンが発症する

①と②が同時に観察されたときのみ、この人においてピロリ菌除菌が20年間胃ガンを発症しないことの原因であると裏付けることができます。

一般に、「対象に原因を与える」場合と「対象に原因を与えない」場合の2つの因果の現象の観察が同一対象で同時に行われていないと、因果関係の有無は判断できません³⁾。表1に今回の例での判断の結果をまとめておきます。

さて、現実を考えてみると、残念ながら1人の人については、「ピロリ菌除菌が行われる」場合を観察してしまうと「ピロリ菌除菌が行われない」場合は観察できないし、逆もそうです。1人の人についてこれらの2つの場合を同時に観察することができないため、この人についてピロリ菌除菌が20年間胃ガンを発症しないことの原因であることを裏付けることはできません。しかしながら、このように現実的には観察不能な2つの因果の現象を観察できたとして、因果関係の有無を議論していくことは、議論そのものの明快な整

表1. 因果関係の有無の判断

		ピロリ菌除菌が行われない	
		胃ガンを未発症	胃ガンを発症
ピロリ菌除菌が行われる	胃ガンを未発症	因果関係なし	因果関係あり
	胃ガンを発症	因果関係なし	因果関係なし

理に役立ちます。このような形で因果の議論を進めていくことを、カウンターファクチュアルモデル⁴⁾を用いた議論と呼びます。

3.2. 集団について考えてみる

前述の通り、1対象について、現象と要因の因果関係を裏付けることは現実的に不可能です。では、集団を対象にして考えてみてはどうでしょう。つまり、ある集団（ターゲット集団）について、「全員にピロリ菌除菌が行われる」場合と「全員にピロリ菌除菌が行われない」場合の胃ガンの発症の様子を同時に観察することを考えてみます。

相変わらず、これら2つの因果の現象を同時に観察するのは現実的に不可能です。しかしながら、前節の標本調査の話の思い出してください。ターゲット集団からランダムサンプリングによってつくられた標本を調べることで、ターゲット集団全員を調べる場合に得られる結果を推測することができます。したがって、ターゲット集団から2回のランダムサンプリングによって2つの標本をつくるのであれば、一方の標本には全員にピロリ菌除菌を行い、他方には全員にピロリ菌除菌を行わず、それぞれの標本の胃ガンの発症の様子を観察することで、ピロリ菌除菌を行う標本からは、ターゲット集団について「全員にピロリ菌除菌が行われる」場合の胃ガンの発症の様子を、投与しなかった標本からは「全員にピロリ菌除菌が行われない」場合の胃ガンの発症の様子をそれぞれ推測することができます。それら2つの推測結果の比較は、まさにターゲット集団に対してカウンターファクチュアルモデルの状況を実現しており、ピロリ菌除菌と20年間胃ガンが発症しないことの因果関係を裏付けることができます。ピロリ菌除菌が行われる標本と行われない標本のそれぞれにおいて胃ガンの発症の様子を20年間観察し、ピロリ菌除菌が行われる標本からは、「ターゲット集団全員にピロリ菌除菌が行われる」場合の胃ガンを発症する人の割合を、行われない標本からは「ターゲット集団全員にピロリ菌除菌が行われない」場合の胃ガンを発症する患者の割合をそれぞれ推測し、それらの推測結果を比較することで、ピロリ菌除菌が胃ガンの発症割合を変化させているかどうかで因果関係を裏付けます。1対象について不可能であったカウンターファクチュアルモデルの状況も、対象を集団に拡張し、標本調査の考えを利用することで実現できるわけです（図2を参照）。

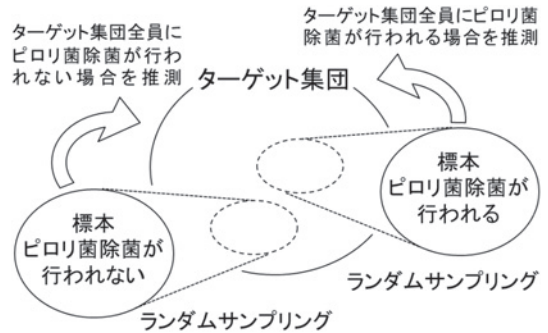


図2. 2回のランダムサンプリングに基づいた比較

3.3. ランダム化比較研究

臨床の現場では、前節で示した考え方を基本にして、次に示すような形で研究が行われることとなります。まず、研究に参加する対象を集め研究参加集団を作ります。次に、ピロリ菌除菌を行うグループと行わないグループに各対象をランダムに割り当てていきま（現実には倫理的に問題があるかもしれませんが、あくまでも例題ということ）。このランダムな割り当てをランダムアロケーション（無作為割り付け）と呼びます⁵⁾。

いま、ピロリ菌除菌を行うグループに割り当てられた1人について考えてみましょう。ランダムアロケーションによってこのグループの一員になったこの対象は、見方を変えれば、研究参加集団からランダムな手続きを経てこのグループの一員として選ばれたとみることができます。つまり、ピロリ菌除菌を行うグループは、研究参加集団からランダムサンプリングによってつくられた標本とみなすことができます。もちろん、ピロリ菌除菌を行わないグループも同様です。したがって、ピロリ菌除菌を行うグループからは、「研究参加集団全員にピロリ菌除菌を行うとき」の胃ガンの発症の様子を推測できることになり、ピロリ菌除菌を行わないグループからは、「研究参加集団全員にピロリ菌除菌を行わないとき」の胃ガンの発症の様子を推測できることとなります。これら2つの推測結果の比較は、研究参加集団についてカウンターファクチュアルモデルを実現しており、これによりピロリ菌除菌が20年間胃ガンを発症しないことの原因であるかどうかを判断できます（図3を参照）。このように研究のために集めた対象をランダムに複数のグループに振り分け、各グループに異なる介入を行って、結果の発症の様子を比較するスタイルの研究をランダム化比較研究（試験）と呼びます。ランダム化比較研究は因果

推論において良質のエビデンスを与える研究デザインとして知られています。

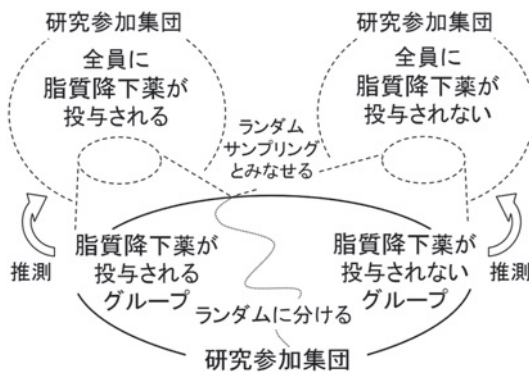


図3. ランダム化比較研究の模式図

4. 検定の考え方

前節では、因果推論の基本になる考え方と因果推論のために必要なデータ収集の理想的な方法（研究デザイン）について解説しました。本節では、集めたデータを分析する際に用いる統計手法について解説します。

4.1. 記述統計手法と推測統計手法

統計手法は大きく記述統計手法と推測統計手法に分けられます。記述統計手法は研究で収集したデータを用いて、データを与えてくれた集団の様子を表現するための手法です。全数調査においては、ターゲット集団全員からデータが収集されるので、記述統計手法によってターゲット集団の様子を表現できます。標本調査においては、記述統計手法で標本の様子のみを表現でき、ターゲット集団の様子は表現できません。そこで、推測統計手法を用いて、標本のデータからターゲット集団の様子を表現します。

このように、記述統計手法は全数調査、標本調査の両方に使用しますが、推測統計手法は基本的に標本調査のみに使用します。標本調査においては、それぞれの名前が示すように記述、推測と役割が異なり、分析は以下の順に進めていくのが基本になります。

- ①標本の様子を記述
- ②ターゲット集団の様子を推測

記述統計手法に関する具体的な手法の紹介等に関しては他書^{6)~9)}に譲ります。

推測統計手法は大きく推定と検定の2つのアプローチに分けられます。推定は、ターゲット集団全員を調べたときに求まる様々な数値を、標本からのデータを用いて、具体的な値として求めようとするアプローチ

です。検定は、ターゲット集団全員を調べたときにわかることについて仮説を立て、この仮説が誤っているかどうかを標本からのデータを用いて判定しようとするアプローチです。

検定はいろいろな場面で頻繁に用いられる推測のアプローチですが、考え方が特異なせいか、その結果が誤用されることも少なくありません。ここでは、因果推論を例に検定のロジックについてわかりやすく解説します。まずは、具体的な検定の解説に入る前の準備として「背理法」、「確率」の2つについて説明をしておきます。

4.2. 背理法（準備その1）

まずは、観察結果に基づいて事前に設定した仮説を否定することで知見を得る背理法についてその手続きを解説します。背理法では、以下に示す手続きを踏みます。

- ①仮説を立てる
- ②観察を行う
- ③観察結果と仮説が矛盾していないか調べる
- ④矛盾していたら仮説を否定する

矛盾していない場合は仮説に対する判断を保留にする

検定ではこの背理法の手続きを踏みます。

4.3. 確率（準備その2）

続いて確率について、検定を理解する上で最低限押さえてほしい2つの話をしておきます。

まずは確率の計算には仮定が必要になるという話です。たとえば、さいころを振ったとき1の目が出る確率はどれくらいでしょう？おそらく、多くの人はその確率を1/6と考えるでしょう。その考えは、「サイコロの6つのどの目も均等に出る」という仮定を暗黙のうちに想定しています。確率計算の裏にはこうした仮定が存在することを意識しておいてください。仮定なくして確率は計算できないことをここで強調しておきます。さらに言えば、仮定を別の仮定に変えると、結果として得られる確率も変化します。確率というのは絶対的に不変な数値ではないということも意識しておくといよいでしょう。

次は、確率を使った判断についての話です。多くの人は日常生活において、無意識下で正確ではないまでも確率を計算し物事の判断に利用しています。たとえば「宝くじで7億円（1等+前後賞）が当たった」という話を聞くと、たいていその話を最初は信じないものです。「7億円はそう当たるものではない」つまり「7億円が当たる確率は非常に小さい」と考え、そ

のことが起きたことを疑うのです。起こる確率が非常に小さい出来事が起こったということに対して、多くの人はそれを疑ったり信じなかったりするものです。こうした起こる確率が小さい出来事に対する感覚や反応が、検定を理解する上で非常に重要になることを強調しておきます。

確率についての2つの話をまとめておきます。

- 1) 確率の計算には仮定が必要
- 2) 起こる確率が小さいと思われる出来事が起きたことに疑いや不信感を持つことが大切

4.4. 検定の考え方

2つの準備を踏まえて、ランダム化比較研究での検定の利用を例に検定の考え方を説明します。まず、検定は背理法の手続きにしたがっています。最初に、研究参加集団全員を調べたときの様子について仮説を立てます。この興味の対象集団について立てる仮説を帰無仮説と呼びます。ランダム化比較研究においては「研究参加集団全員にピロリ菌除菌を行う場合と行わない場合とで、20年間で胃ガンを発症する人の割合は変わらない」という帰無仮説を立てます。次に、研究参加集団の観察を行います。これは、研究参加集団を、ピロリ菌除菌を行うグループと行わないグループの2つにランダムに分けて胃ガンの発症についてデータを採ることに対応します。したがって、2つのグループのそれぞれから採られたデータが観察結果ということになります。続いて、観察結果であるデータが帰無仮説と矛盾しないかを判断します。この判断のために次

の確率を求めます。

「研究参加集団全員を調べたときの様子が帰無仮説の様な状態にある（研究参加集団全員にピロリ菌除菌を行う場合と行わない場合とで、20年間で胃ガンを発症する人の割合は変わらない）とした場合の、今回の研究で採られたデータの出現確率（のようなもの）」

では、この確率がそれなりに大きい値をとったときを考えてみましょう。この場合、データの出現する確率が大きい値となったわけですから、その出現に疑いを持たない、つまり、帰無仮説のような状況からデータは出現してもおかしくないということになります。このことはデータと帰無仮説の間に矛盾がないことを示しています。逆に、確率が小さい値になった場合は、データの出現する確率が小さい値となったわけですから、データの出現に疑いを持つ、つまり、帰無仮説のような状況からデータが出現するなんて信じられない（疑わしい）ということになります。このことはデータと帰無仮説の間に矛盾があることを示しています。このように上記の出現確率を計算することで、データと帰無仮説の間に矛盾があるかどうかを判定できます。最後に、背理法の手順にしたがい、矛盾があった場合は帰無仮説を否定し、矛盾がなかった場合には、帰無仮説に対する判断を保留します。矛盾がなかった場合の判断には注意してください。決して帰無仮説が正しいという判断にはならないことを強調しておきます。

データと帰無仮説が矛盾するかどうかの判断に用いたこの確率のことを p 値と呼びます。 p 値が小さい

表2. 検定のまとめ

<背理法>	<検定>
① 仮説を立てる	1) 帰無仮説を立てる 興味の対象集団全員を調べたときの様子に関して立てる
② 観察する	2) 研究参加集団を観察する 興味の対象集団よりランダムサンプリングにて抽出した標本からデータを採る
③ 観察結果と仮説が矛盾していないか調べる	3) データと帰無仮説が矛盾していないか調べる 興味の対象集団全員を調べたときの様子が帰無仮説の様な状態にあるとした場合の、標本から採られた手元のデータの出現確率（のようなもの）を計算（ p 値を計算） ・ p 値が有意水準より大きい場合 「帰無仮説のような状況からデータは出現してもおかしくない」となり、データと帰無仮説の間に矛盾はない ・ p 値が有意水準より小さい場合 「帰無仮説の状況からデータが出現するなんて信じられない（疑わしい）」となり、データと帰無仮説の間に矛盾が生じていることになる
④-1 矛盾していた場合 仮説を否定	4-1) p 値が有意水準より小さい場合 帰無仮説を否定
④-2 矛盾していなかった場合 仮説に関する判断を保留	4-2) p 値が有意水準より大きい場合 帰無仮説に対する判断を保留

値かどうか（矛盾があるかどうか）の判断には前もって基準を用意しておく必要があります。この判断基準を有意水準と呼びます。有意水準には0.05（5%）という値が一般に用いられますが、状況に応じて事前に決める値で、必ず5%にしなければならないということはありません。表2にこれまでの話を背理法と対比して整理しておきますので、参照してください。

5. データの尺度と検定手法

本節では、因果推論を目的とした研究、特にランダム化比較研究においてよく用いられる検定手法を紹介いたします。検定手法の選択においては、原因の項目と結果の項目がどのような尺度で測定されているかが鍵となります。そこで、まず尺度の概念について説明をした後で、尺度の組み合わせごとに検定手法を紹介します。なお、各検定の計算方法などの詳細については他書^{9)~8)10)}に譲ります。

統計解析が取り扱うデータは、大きく連続尺度データ、順序尺度データ、名義尺度データの3つに分類されます。更に細かく分類できるのですが、ここでは検定手法の選択に必要な範囲にとどめておきます。説明を容易にするために身長データの採用を例に挙げます。

5.1 連続尺度データ

身長データと言って思い浮かべるのは「○○○cm」と言った数値のデータでしょう。このようにデータが数値で測定され、かつ、その数値の大小が何かの程度を表している場合（身長の例では、数値が大きいほど身長が高くなるといった身長の程度を表現している）、そのデータを連続尺度データと呼びます。

5.2 順序尺度データ

身長データを、以下に示すような選択肢（カテゴリ）を準備して、該当するカテゴリにチェックをつけさせるスタイルで採ることもできます。

- 160 cm 未満
- 160 cm 以上 170 cm 未満
- 170 cm 以上 180 cm 未満
- 180 cm 以上

回答のために準備されたカテゴリを見ると、明らかにカテゴリ間に順序性（上から下に行くにしたがって背が高くなる）があります。このように、3つ以上の明確な順序性のあるカテゴリから1つを選択するスタイルでデータが採られている場合、そのデータを順序尺度データと呼びます。

5.3 名義尺度データ

身長データを、非常におおざっぱに、以下に示すようなカテゴリを準備して、該当するカテゴリにチェックをつけさせるスタイルで採ることもできます。

- 低い（170 cm 未満）
- 高い（170 cm 以上）

このようにカテゴリが2つしかない場合に限らず、明確な順序性がないカテゴリから1つを選択するスタイルでデータが採られている場合、そのデータを名義尺度データと呼びます。

順序尺度データと名義尺度データはいずれも測定のためにカテゴリが準備されるので、2つの尺度を併せてカテゴリカルデータ（離散データ）と呼びます。カテゴリカルデータの内、順序尺度データでないものが名義尺度データということになります。

5.4. 検定手法の使い分け

因果推論では、研究において原因と考えている項目と結果と考えている項目の2つの項目が存在します。それぞれの項目がどのような尺度のデータであるかによって、用いる検定手法が変わってきます。表3に各尺度の組み合わせで用いられる代表的な検定方法を挙げておきます。

たとえば、原因の項目が名義尺度データで、結果の項目が連続尺度データの場合、2標本t検定（原因の

表3. 因果推論に用いられる検定手法

		結果の項目		
		連続尺度	順序尺度	名義尺度
原因の項目	名義尺度	<ul style="list-style-type: none"> ・ 2標本t検定 ・ 1元配置分散分析 ・ U検定 ・ Kruskal-Wallis検定 	<ul style="list-style-type: none"> ・ U検定 ・ Kruskal-Wallis検定 ・ CMH検定* 	<ul style="list-style-type: none"> ・ χ^2検定 ・ Fisherのexact検定
	順序尺度	<ul style="list-style-type: none"> ・ 分散分析 	<ul style="list-style-type: none"> ・ CMH検定* 	<ul style="list-style-type: none"> ・ CMH検定*
	連続尺度	<ul style="list-style-type: none"> ・ 線形回帰分析 	<ul style="list-style-type: none"> ・ 順序ロジスティック回帰分析 	<ul style="list-style-type: none"> ・ ロジスティック回帰分析

*CMH検定：Cochran-Mantel-Haenszel検定

項目が曝露・非曝露のような2カテゴリのとき)や1元配置分散分析(原因の項目が3カテゴリ以上のとき)などを用いることになります。いずれの検定も「各ターゲット集団の結果の項目の平均値は等しい」を帰無仮説としています。

ランダム化比較研究では、研究参加集団を複数のグループにランダムに分け、各グループに原因の項目のカテゴリを割り当てていきます。割り当てたカテゴリの個数や順序性の有無によって、原因の項目は順序尺度データもしくは名義尺度データのいずれかになります。たとえば、2つのグループを作成する場合は、原因の項目のカテゴリ数が2となるので、その項目は名義尺度データということになります。3つのグループを作成し、各グループに低用量、中用量、高用量のように、異なる用量の曝露を割り当てた場合は、原因の項目は順序尺度データということになります。あとは結果の項目の尺度に応じて用いる検定を決めることになります。いずれの検定も「研究参加集団において、原因の項目と結果の項目の間に関連はない」といった意味に解釈できる帰無仮説が設定されています。たとえば、前出のピロリ菌除菌と胃ガンのランダム化比較研究の場合、「研究参加集団全員にピロリ菌除菌を行う場合と行わない場合とで、20年間で胃ガンを発症する人の割合は変わらない」という帰無仮説を設定する χ^2 検定(原因の項目、結果の項目ともに名義尺度データなので)を用います。この帰無仮説は、言うまでもなく、「研究参加集団においてピロリ菌除菌と胃ガンの発生には関係がない」と解釈できます。

6. 最後に

本稿では、因果推論を中心にそれに必要な統計学の知識の解説と理想的な研究デザインの紹介などを行いました。残念ながら、すべての因果の問題において、本稿で紹介したランダム化比較研究を行うことはできません。ランダムに原因の項目を対象に割り当てた

のが医療倫理や研究倫理に抵触する可能性があるからです。その場合、ランダムに割り当ててそのものをあきらめる、もしくは原因の項目の割り当てそのものをあきらめることになります。特に後者の場合は、代わりに観察研究と呼ばれる研究を行うことになるかもしれません。いずれの場合も、原因の項目をランダムに割り当てないことにより生じる交絡バイアスと呼ばれるやっかいな問題に対処する必要があります。本稿においてこの問題についても解説をしたかったのですが、紙面の都合上割愛させていただきました。もし、次の機会があれば、観察研究やバイアスの問題などについて解説したいと思っております。興味のある方は、他書⁹⁾¹¹⁾の解説を参照することをお勧めいたします。

文 献

- 1) 椿美智子, 椿 広計. 2001. 医学研究のための統計的方法, サイエンス社.
- 2) Everitt, B. S. 2006. 医学統計学辞典(宮原英夫, 池田憲昭, 鶴田陽和監修), 朝倉書店.
- 3) 佐藤俊哉. 1999. Pコントロール, p. 21-34, これからの臨床試験(椿 広計, 他編), 朝倉書店.
- 4) Greenland, S., B. Brumback. 2002. An overview of relations among causal modelling methods. *International Journal of Epidemiology* 31: 1030-1037.
- 5) 西 次男. 2006. 無作為化, p. 137-155, 臨床試験ハンドブック(丹後俊郎, 他編), 朝倉書店.
- 6) 折笠秀樹監訳: 2003. はじめて学ぶ医療統計学, 総合医学社.
- 7) 藤井良宜. 2009. 統計学—その基本的な考え方—, 放送大学教育振興会.
- 8) 栗原伸一. 2011. 入門統計学, オーム社.
- 9) 比江島欣慎. 2014. 医療統計学入門, オーム社.
- 10) 柳川 堯. 1982. ノンパラメトリック法, 倍風館.
- 11) Rothman, K. J. 2004. ロスマンの疫学(矢野栄二, 橋本英樹(監訳), 篠原出版新社).