

特集 I 論文を書こう！

研究デザインとエビデンス

総説▶

松尾浩一郎

Koichiro Matsuo

要旨 質の高い臨床研究を行うためには、まず検証しやすい研究仮説を立て、その仮説を実行しやすい研究デザインのもとで行う必要がある。また、研究結果を間違いなく解釈する力も必要となる。本稿では、研究仮説の立て方から各種の研究デザイン、研究で陥りやすいピットフォール、エビデンスなどについて解説していく。

キーワード▶ 研究仮説、研究デザイン、エビデンス、バイアス、交絡因子

I. 研究デザイン

研究論文を書くためにはまず研究をしなくてはならない。そして、質の高い臨床研究を行うためには、検証しやすい研究仮説を立て、実行しやすい研究計画を立てることが重要である。また、研究結果を間違いなく解釈する必要もある。簡潔で具体的な仮説のもとで研究を行うと論文も書きやすくなる。本項では、研究仮説の立て方から各種の研究デザイン、研究で陥りやすいピットフォールについても解説していきたいと思う。

臨床研究には、研究起案から最終的な解釈まで決まった一連の流れがある（図 1）。一般的な臨床的疑問を検証するために、ある具体的な研究仮説を立て、その仮説のもとに研究計画をデザインする。その実験条件のなかで統計学的に有意な差を見いだしたならば、その結果が自分の疑問点（仮説）にどのように答えているかを推論する。その答えから、最終的に一般的な臨床問題へどのように帰属すべきか導いていく。この流れを理解し咀嚼すると、臨床研究を行いやすくなる。

II. 研究仮説の立て方

1. 良い仮説の条件

研究を開始する前に研究仮説を立てなければならない。臨床家の場合、日常臨床の疑問の検証から始まることが多いのではないだろうか。しかし、どんなにいい Clinical Question であったとしても、それを実験系にうまく落とし込まなければ質の高い研究はできない。そこでまず大事になってくるのが、研究仮説を立てることである。反対に、

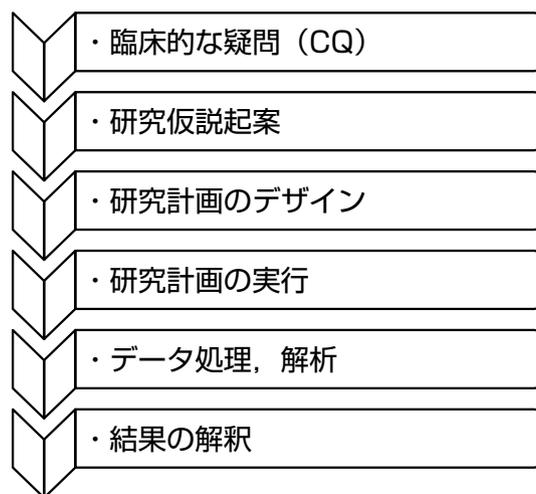


図 1 研究の流れ

研究仮説を立てずに研究を始めてしまうことほど残念なことではない。Fishing Experiment という言葉をご存じだろうか。Fishing Experiment とは、具体的な仮説をもつことなしに、いろいろな項目を調べて、有意な値がでるまで分析を行う研究のことである。基本的に P 値=0.05 を有意水準とすると、20 項目の統計分析を行うと、そのうちの 1 項目は、たとえそこに因果関係がなくても、偶然に統計学的有意な差がでる理屈になる。そこに統計学的有意な関連性があると結論づけたとしても、そこに因果関係を見いだすのは難しく、結果の解釈に悩むことになる。研究は釣りではない。原因と結果の関係を明らかにするためには、釣りを行うのではなく、まず初めに具体的な研究仮説を立て、その仮説の検証を行うことが重要となる。それが仮説誘導型の研究 (Hypothesis Driven) と呼ばれる形

藤田保健衛生大学医学部歯科

[連絡先] 松尾浩一郎：藤田保健衛生大学医学部歯科 (〒470-1192 愛知県豊明市沓掛町田楽ヶ窪 1-98)

TEL : 0562-93-9097 E-mail : kmatsuo@fujita-hu.ac.jp

受稿日：2015 年 9 月 8 日 受理日：2015 年 9 月 11 日

- ・ Specific (具体的に)
- ・ Simple (簡潔に)
- ・ In advance (研究を始める前に)

図2 良い仮説の条件

である。

臨床研究の始まりの多くは、臨床で感じた疑問点 (Clinical Question, CQ) から始まるのではないだろうか。このCQを研究できるような疑問点 (Research Question, RQ) に置換するところがポイントとなる。そのときに、いかにシンプルな研究仮説へと変換することができるかが、後の研究デザインに大きく影響する。臨床的な疑問点から研究仮説へと変換することで、初めてその研究計画がデザインできる。仮説を立てることは、論文作成に限ったことでない。研究仮説は、研究費の申請書のなかでもキーとなる部分である。

それでは、良い研究仮説とはどのようなものなのか。アメリカのジョンズホプキンス大学公衆衛生学部の疫学講座の元教授であるDr. Jonathan M. Sametが、「A hypothesis is a statement that can be tested (仮説とは検証できる一文である)」と述べている。逆に言うと、仮説は実験で検証できなくてはならない。良い研究仮説とは、具体的に (Specific)、簡潔に (Simple)、かつ研究を始める前に (In advance) 立ててあるものである¹⁾ (図2)。次に、良い研究仮説の骨格について説明していく。

2. 仮説の骨格

研究仮説には骨格がある (表1)。その骨格は、後述する研究のデザインによって変わるのだが、PECOやPICOなどの組み合わせがよく使われる。対象者の疾患危険因子と疾患発症との関連性を検討するような観察研究の場合には、PECOが用いられ、治療による臨床的帰結との関連性を検討する介入研究ではPICOが用いられることが多い。PECO、PICOは、仮説の骨格のそれぞれの頭文字である。PECOは、Patients, Participants (患者, 対象者) に対して、どのようなExposure (曝露: 治療, 危険因子) があると、その曝露がなかった場合とComparison (比較) して、Outcome (結果: 疾患, 臨床的帰結) がどのようになるか。という要素になる。介入研究の場合には、ExposureがIntervention (介入: 治療など) となる。これらの項目を骨格とすることで、仮説を立てやすくなる。

今回は、一つ例を挙げて説明していく。摂食嚥下障害への有名な筋力訓練の一つに頭部挙上訓練、通称シャキアー訓練がある。Dr. Shakerによって1997年に開発されたシャキアー訓練は嚥下訓練のなかでも比較的高いエビデン

表1 仮説の骨格

観察研究: PECO

Patients, Participants (患者, 対象者)

Exposure (曝露: 治療, 危険因子)

Comparison (比較)

Outcome (結果: 疾患, 臨床的帰結)

介入研究: PICO

Patients, Participants (患者, 対象者)

Intervention (介入: 治療など)

Comparison (比較)

Outcome (結果: 治療成績, 臨床的帰結)

スが示されている訓練である^{2,3)}。このシャキアー訓練の有用性について報告された研究をみながら、研究仮説を説明していく。

CQから研究仮説を立てるにいたった経緯としては以下のような流れとなる。嚥下障害者をVFで評価したら、梨状窩に多量の残留を認めた。食道入口部 (UES) の開大不全が、その残留の原因と考えられた。しかし、UESの開大不全に対して、効果的な訓練方法はそれまでに確立されていなかった。そこで、彼らはUESの開大改善のための新たな訓練方法を提案した。UES開大のメカニズムは、以下の3つの機序が関与すると考えられている。(1) 喉頭 (輪状軟骨) を起始とする輪状咽頭筋が弛緩する。(2) 同時に、舌骨上筋群の収縮により、舌骨、喉頭が前上方に挙上することで、輪状咽頭筋が前方に牽引され、UESが開かれる。(3) また、咽頭から送り込まれてきた食塊によってさらにUESが開大する。Shakerらは、このうち(2)の舌骨上筋群の収縮による喉頭挙上に注目した。彼らの立てた研究仮説は、摂食嚥下障害者 (P) において、舌骨筋群の増強訓練によって (I)、シャム訓練 (コントロールとしての偽の訓練) よりも (C) 食道入口部の開大量が増大する (O)、またそれによって、咽頭残留が軽減する (O)、であった。仮説を立てるときのポイントは、シンプルに具体的にである。「Shaker訓練により嚥下障害が良くなる」よりも、「Shaker訓練により食道入口部の開大が改善する」のほうがより具体的であり、実験系が組み立てやすくなり、結果の解釈が容易となる。

Ⅲ. 研究計画の種類と注意点

研究仮説を立てた後に、その仮説を検証するために研究計画を立案する。研究のデザインにはいくつかあるが、研究規模、期間、費用、人手、などいろいろな要因を検討したうえで、どのような研究デザインにするか決定する。この研究デザインによって、研究成果のエビデンスとしての

表2 研究デザインの種類

観察研究 Observational Study
横断研究 Cross-Sectional Study
ケースコントロール研究 Case-Control Study
コホート研究 Cohort Study
介入研究 Intervention Study
非ランダム化比較試験 Non-Randomized Controlled Trial
ランダム化比較試験 Randomized Controlled Trial

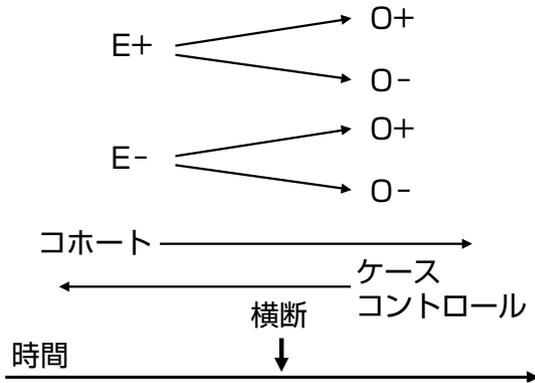


図3 観察研究における時間的關係。横断研究はある一時点で行う研究。ケースコントロールは、時間軸を後ろ向きに検討し、コホート研究は、前向きに検討していく。

レベルが決まる。研究デザインは、まず研究対象に対して介入を行うかどうかによって、介入する介入研究と介入しないで観察するのみの観察研究とに大別される⁴⁾(表2)。介入研究は、対象を無作為に群別化して実施されるランダム化比較試験と非ランダム化比較試験に分けられる。観察研究は、さらに研究の曝露と帰結の時期の設定や、実験の時間的方向性(前向きか後ろ向きか)などによって、横断研究、ケースコントロール研究、コホート研究などに分けられる(図3)。

1. 観察研究 (Observational Study)

1) 横断研究 (Cross-Sectional Study)

横断研究では、ある一時点での曝露の有無 (E+/-) と結果、疾病の有無 (O+/-) を観察し、両者の関連性について分析する。横断研究の利点は、研究期間が短く、費用もそれほどかからないことである。一方で、その大きな欠点として、E と O との両者の因果関係は明らかにはならないことが挙げられる。因果関係を明らかにするには、E が O より時間的に先行しなければならない。しかし、横断研究ではある一時点での観察であるので、両者の時間的關係はわからない。つまり、たとえ横断研究で、E と O との間に有意な関連性を認めたとしても、E が原因で O となっているのか、または O が要因となって E が引き起

A) 義歯と認知症との関連性を調査

	認知症有	認知症無	計
義歯有	50	150	200
義歯無	200	100	300
計	250	250	500

B) Eを義歯、Oを認知症として考えると、義歯を入れていないと認知症になる人が67%

	認知症有	認知症無	計
義歯有	50 (25%)	150	200
義歯無	200 (67%)	100	300
計	250	250	500

C) Eを認知症、Oを義歯として考えると、認知症の人の80%が義歯を入れていない。

	認知症有	認知症無	計
義歯有	50	150	200
義歯無	200 (80%)	100 (40%)	300
計	250	250	500

図4 横断研究の一例。(A) 義歯と認知症の関連性について調査したと仮定して、両者の因果関係を見てみると、(B) 曝露 (E) を義歯、結果 (O) を認知症として考えると、義歯有りの人が認知症になるのが25%であるのに対して、義歯無しでは認知症になる人が67%とかなりの割合の増加を認める。(C) しかし、Eを認知症、Oを義歯として考えると、認知症の人の80%が義歯を入れていない、というだけのことになる。横断研究の結果から、両者の因果関係について示唆するのは難しい。

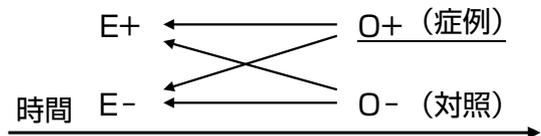


図5 ケースコントロール研究の模式図。まず現時点の症例と対照群 (O+/-) を集めてから、それぞれ過去に何が合ったのかさかのぼってそのリスク因子 (E+/-) を観察する。

こされているかという時間的關係はわからない。そのため、研究結果から E と O の因果関係に結びつけるのは困難である。簡単な例を挙げて説明したいと思う。たとえば、ある老人保健施設で、義歯使用と認知症との関連性を調査したとする。その結果を図4に示す。一見、義歯を使用していないと認知症の割合が多くなるように見えるが、その一方で、ただ単に認知症がある方のほうが義歯をなくしやすかっただけなのかもしれない。横断研究では、両者の時間的な因果関係を明らかにすることはできない。

2) ケースコントロール研究 (Case-Control study)

ケースコントロール研究では、まず現時点での結果の有無、または症例と対照 (O+/-) のデータを集めて、それぞれ過去に曝露があったのか (E+/-) をさかのぼって観

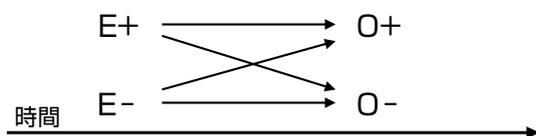


図6 コホート研究の模式図。現時点で、E+とE-の群を集めて、結果がわかるまで(O+/-)まで前向きに調査していく。

察し、両者の差異を分析する(図5)。まず、現時点での症例と対照を集めてから過去にさかのぼっていくので、まれな疾患(O)の調査などに有用である。稀少疾患の場合、前向きコホート研究でその疾患の発症を調査しようとすると、非常に多くのサンプルを集めなければならない、時間がかかってしまう。ケースコントロール研究では、サンプルを症例からピックアップできるので、サンプルを集めやすく、調査期間もコホート調査より短期間で済む。

一方で、ケースコントロール研究の欠点としてバイアスのリスクが挙げられる。バイアスの詳細については後述する。コントロールを選ぶときには、できるだけ症例と同条件のサンプルを選ばなければならないが、そのときに選択バイアス(Selection Bias)が入りやすく、適切なコントロールを選別するのが難しいことが多い。また、アンケート集計などでは、そのときの情報が記憶のなかで変わっている可能性がある。このような情報の偏位を情報バイアス(Information Bias)と呼ぶ。ケースコントロール研究では、研究デザインの段階でも研究データ採取、分析の段階でもバイアスのリスクが入りやすいといわれている。

3) コホート研究(Cohort Study)

コホート研究は、曝露あり(E+)と曝露なし(E-)の群を集めて、結果がわかるまで("O+/-")まで前向きに観察し、EとOの関係について分析する(図6)。リスクあり群となし群で、病気の発症率の違いをみるときに用いられる。コホートとは、研究で追跡していく対象者の集団を意味する。コホート研究では、研究で検証されるEについてOが明らかになる前に測定されるために、Eが時間的にOに先立つことになる。そのために、横断研究よりも、両者の時間的な因果関係をより強く推察できる。また前向き調査では、調べたい調査項目について細かく調査できるメリットもある。ケースコントロール研究では、過去の因子を調査するので、必要なデータが欠損していたり、対象者の記憶が曖昧なためにデータの信頼性が低下することがあるが、コホート研究ではそのようなバイアスを取り除くことができる。まれなEに対する研究に用いやすい。

一方、コホート研究の欠点として、大きいサンプルサイズや長期間の観察期間が必要になることが多い。そのために研究資金も高額になりやすい。またまれな疾患(O)の調査は、膨大なサンプル数が必要になってしまうために不

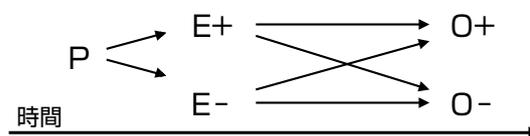


図7 ランダム化比較試験の模式図。対象者(P)に対してランダムにEを与えてE+/-の両群がどのような結果"O+/-"になったかを比較する。

向きである。

2. 介入研究(Intervention Study)

介入研究は、対象を無作為に群別化して実施されるランダム化比較試験と、非ランダム化比較試験に分けられる。両者の違いは割付だけであるので、ここではランダム化比較試験について説明する。

1) ランダム化比較試験(Randomized Control Trial, RCT)

対象者に対して無作為に介入(E)を与えてE+/-の両群がどのような結果"O+/-"になったかを観察し分析を行う(図7)。介入試験のメリットは、観察研究よりもバイアスや交絡因子(後述)の影響を除去しやすいため、EとOとの因果関係を明らかにするためには一番有用な試験である。ただ、時間とコストがかかることや、非倫理的になりやすいことに注意しなければならない。また、通常の試験の場合では介入が一つしか行えないために、CQが狭くなってしまうことが多い。

再びShaker訓練を例としてRCTを説明していきたい。2002年のShakerらの論文である。彼らは「舌骨筋群の増強訓練によって食道入口部の開大量が増大するもしくは咽頭残留が軽減する」という仮説を、経管栄養中のUES開大不全を伴う嚥下障害者を対象に検証した³⁾。対象者をランダムにShaker訓練を行う群とSham(偽の)訓練を行う群の2群に割り付け、6週間の訓練を実施させた。その結果、Shaker訓練群では、喉頭の前方挙上とUESの前後径が有意に増加しており、また、嚥下後の梨状窩の残留も有意に減少したことが示されていた(図8)。嚥下訓練のエビデンスレベルは低いといわれていた当時、初めてRCTで嚥下訓練の効果を検証した研究だったのではないかと思う³⁾。

IV. 陥りやすい落とし穴

1. EとOとの関係性

臨床研究の結果として、特に観察研究において、E(曝露)とO(結果)との間に有意な関連性が認められたときに、両者の間に因果関係があると結論づけたいところだが、そこには注意が必要である。関連があるのと因果関係があるのでは意味が異なる。両者の間に関連性が認められたとき

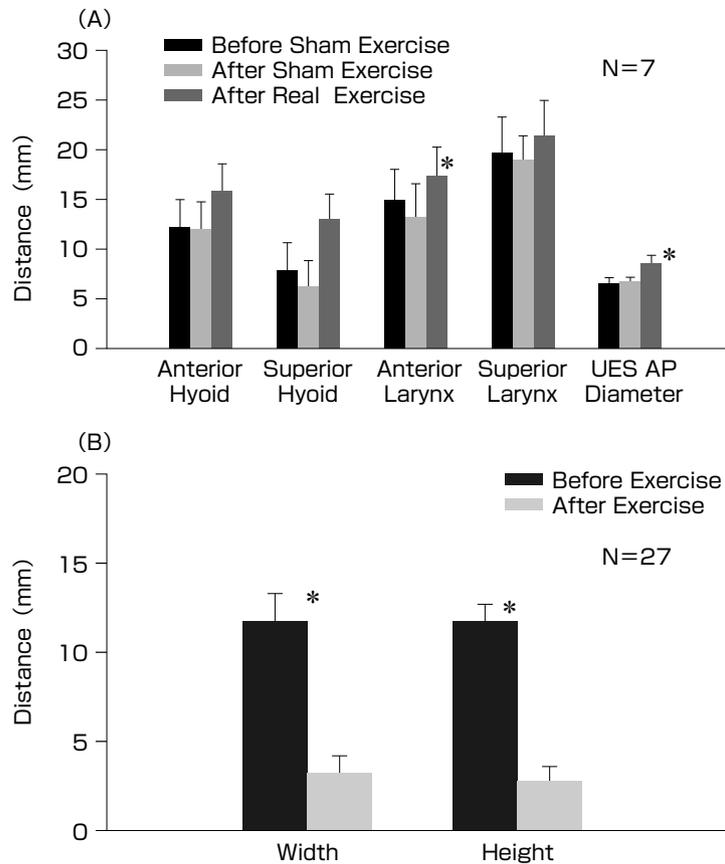


図8 Shaker et al, 2002 より抜粋³⁾. (A) Sham 群と比較して Shaker 訓練群は、喉頭の前方移動、UES の前後径において訓練前からの有意な増加を認めている。(B) Shaker 訓練群では、訓練前よりも有意に梨状窩残留の幅も高さも減少している。* : $p < 0.05$

表3 両者の関係¹⁾

1. 偶然誤差 Chance (Random Error)
2. バイアス Bias (Systematic Error)
3. O → E 効果-原因 関係
4. 交絡因子 (Confounding Factor)
5. E → O 原因-効果 関係

には、両者の関係性にはいくつかの可能性がある。ここでは、真の因果関係以外にどのような可能性があるのか説明していく。

E と O との関係は、大きく5つに分けられる(表3)¹⁾。偶然誤差 (Chance [Random Error]) とバイアス (Bias, [Systematic Error]) は、偽性相関 (Spurious Association) と呼ばれる偽の関連性である。両者の間には真の関連性はなく、研究結果としてそのような関連性がでてしまった関係のことである。また、両者の間に真の関連性 (Real Association) があつたとしても、それが E → O の因果関係であるとは限らず、実は O が原因で E が結果であることもあるし、両者の間に真の O の原因因子であり、

かつ、E とも関連性を有する交絡因子 (Confounding Factor) が存在することもある。以上のように、真の因果関係以外の関係を常に考慮しなければならない。

2. 偶然誤差 (Chance [Random Error])

本来両者には関連性がないにもかかわらず、研究サンプルで有意な関連性がでてしまうのが、ランダムエラーである。ランダムエラーは、ある一定の割合で起こってしまうエラーである。ランダムエラーは、第1種の過誤 (Type I Error, α) と呼ばれ、統計分析で棄却域として設定される p (Probability) = 0.05 と大いに関係している。研究のなかで、E と O との間に偶然にも有意な関係が出てしまうことがある。これが、Type I Error である。この偶然の可能性は0にはできないため、どの程度までだったら認めるかを決めなければならない。その設定で頻用されるのが、5%である。5%以下で発生してしまうランダムエラーは仕方がないという設定である。ランダムエラーを避けるためには、サンプルサイズを大きくする、棄却する P 値を厳しく設定するなどの方法をとる必要がある。

3. バイアス (Bias [Systematic Error])

バイアスとは、研究結果に偏りを生じさせてしまう要因

表4 エビデンスのレベル⁵⁾

I	: システマティックレビュー/RCT のメタアナリシス
II	: 1 つ以上の RCT による
III	: 非 RCT による
IVa	: コホート研究による
IVb	: 症例対照研究, 横断研究による
V	: 記述研究 (症例報告や症例集積) による
VI	: 患者データに基づかない, 専門委員会や専門家個人の意見

のことであり, 系統誤差 (Systematic Error) とも呼ばれる。偶然誤差とは異なり, サンプルサイズを大きくしてもそのリスクは変わらない。バイアスは, 研究デザインからデータ解析の段階まで起こる可能性があり, 選択バイアス (Selection Bias) と情報バイアス (Information Bias) に分けられる。交絡 (Confounding) もバイアスの一つなのだが, ほかの二つとは独立して, あとで説明する。選択バイアスは, 研究デザインで起こりやすく, 情報バイアスは, データ収集から解析の段階で起こりやすい。

選択バイアスは, 研究サンプルがその母集団を正しく代表せずに偏りが起こってしまうことである。実験系の被験者や事象が本来検証すべき母集団を代表していないと, 実験結果から誤った答えを導いてしまう。たとえば, アンケート調査で, 曝露や疾患がある人の回答率が低くなったり, 逆に積極的に回答する場合がある。このような場合には, 対象者の偏りが出てしまい, 対象者が母集団からランダムに抽出されたとはいえない。

情報バイアスとは, データを測定, 収集する方法が結果の偏りを引き起こすものである。白衣高血圧は典型的な情報バイアスである。血圧を測定するときに, 病院で血圧測定すると自宅で測定している値よりも高い値が出てしまう。これは, データ測定での偏りである。また, ケースコントロール研究で, 過去の情報を思い出すときなどにも起こりやすい。実際疾患に罹患した人のほうが, そうでないコントロールの人よりも, 過去の曝露について鮮明に覚えていることが多いためである。

4. 交絡因子 (Confounding Factor)

交絡因子とは, O の原因であり, かつ E と関連性をもつ因子である。交絡因子が存在すると, 表面上は E が O に影響を及ぼしているようにみえるのだが, 実際 O に影響しているのは交絡因子 X である。簡単な例だと, コーヒー (E) と脳梗塞 (O) の間に因果関係があるようにみえるが, たばこを吸う人はコーヒーをよく飲むので, 実は交絡因子であるたばこ (X) が脳梗塞の原因であることがわかる。

研究デザインの段階で交絡因子の影響を取り除くための手段として, 特異化 (Specification) がある。特異化は, 想定される交絡因子を決定し, その条件を対象者から外し

表5 推奨グレード⁵⁾

A	: 強い科学的根拠があり, 行うよう強く勧められる
B	: 科学的根拠があり, 行うよう勧められる
C1	: 科学的根拠はないが, 行うよう勧められる
C2	: 科学的根拠がなく, 行わないよう勧められる
D	: 無効性あるいは害を示す科学的根拠があり, 行わないよう勧められる

てしまう方法である。上のコーヒーと脳梗塞の例だと, 対象者をすべて非喫煙者で集めてしまう。しかし, 特異化のデメリットとして, 喫煙者での影響がわからなくなるため, 研究結果を一般化しにくくなってしまう。一方, 解析の段階で, 交絡因子の影響を取り除くために用いられるのが, 比較する群間で交絡因子のレベルをそろえる層別化 (Stratification) や, 多変量解析で交絡因子も一緒に独立変数に入れてその影響を取り除くことができる補正 (Adjustment) などがある。

V. エビデンス

1. エビデンスレベル

診療ガイドラインが策定される場合には, PECO または PICO をベースとした CQ を設定して, 各 CQ に対して科学的根拠に基づいて (Evidence Based Medicine) 解を得ていくのが一般的である。解を得るときの信頼度を示す指標がエビデンスレベルであり, エビデンスレベルが高いほど信頼性が高いといえる。ガイドライン作成にあたっては, 専門の委員会によって CQ に関連した論文が集められ, 前述した研究デザインなどによって, それらの文献のエビデンスレベルが決定される (表4)。個々の文献を評価するときには, 研究のバイアスリスクなども評価される。介入研究のほうが観察研究よりエビデンスレベルが高く設定されており, RCT をいくつか集めて分析されたメタアナリシスのエビデンスレベルが一番高い。ガイドラインの作成, エビデンスレベルについては, 医療情報サービス Minds (マインズ) のホームページに詳しく掲載されている⁵⁾。ガイドライン作成の手引きの PDF がホームページよりダウンロードできるようになっているので, 興味がある方は参照されたい。

2. 推奨グレード

ガイドライン策定では, 集められた研究のエビデンスレベルに準じて, 診療ガイドラインにおける CQ に対する推奨グレードが決められる (表5)⁵⁾。推奨グレードは, 「強く勧められる」から「行わないよう勧められる」までの5段階に分けられることが多い。一般的に, エビデンスレベルが高いほど, 推奨グレードも高くなる。

VI. ま と め

今回は、研究仮説の立て方、研究デザインの形、エビデンスなどについて概説してきた。詳しいことは、参考文献にある成書^{1, 4)}を参考にしていきたい。論文作成はトライアンドエラーの繰り返しである。普段の臨床的疑問を臨床研究につなげるという作業は大変だと思うが、本項が少しでも読者各位の論文作成の参考になれば幸いである。

文 献

- 1) Hulley SB, Cummings SR, Browner WS, et al : Designing clinical research. Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2013.
- 2) Shaker R, Kern M, Bardan E, et al : Augmentation of deglutitive upper esophageal sphincter opening in the elderly by exercise. *Am J Physiol*, 272 : G1518–1522, 1997.
- 3) Shaker R, Easterling C, Kern M, et al : Rehabilitation of swallowing by exercise in tube-fed patients with pharyngeal dysphagia secondary to abnormal UES opening. *Gastroenterology*, 122 : 1314–1321, 2002.
- 4) 福原俊一：臨床研究の道標：7つのステップで学ぶ研究デザイン。健康医療評価研究機構，京都市，2013。
- 5) 福井次矢，山口直人，森実敏夫，他：Minds 診療ガイドライン作成の手引き 2014。医学書院，東京，2014。