

医療情報

を見る、

医療情報

から見る

エビデンスと向き合うための

10のスキル

青島周一 著

Review medical information,  
Rethink from medical information:  
10 skills to struggle evidences

## 《 2. 妥当性の高い研究デザインとその限界 》

たとえ集団を対象に比較を行った研究であっても、医学的な介入を受ける前後を比べる手法では「三た論法」が成立してしまう余地があり、Efficacy の検出は困難であることは繰り返し述べてきました。

医学的介入、喫煙や生活習慣などが健康状態にもたらす**因果効果**、すなわち**Efficacy**を知るためには、**個人ではなく集団を対象**とした研究を行うことに加えて**適切な比較対照群**を設定することが肝要なのです。

本項では、Efficacy を検出するための代表的な研究手法として**コホート研究**、**ランダム化比較試験**、**症例対照研究**、そして**メタ分析**という研究手法を紹介します。提示されたエビデンスが、どのような研究手法で行われたデータなのかを知るだけでも、結果の一般化可能性や、結果そのものの妥当性、つまり示された研究結果がどれだけ純粋な Efficacy を示しているかについて、おおよそ把握することができます。

### 1. コホート研究

**コホート**とは、一定期間にわたり追跡調査を受ける集団のことで、共通の要因を持った個人の集合体を指します。もともとは古代ローマの歩兵隊の一単位で、数百人からなる兵隊の一群を意味する言葉でした。コホートに登録されている患者集団のデータを用いて、健康状態と**曝露**因子との関連性を一定期間にわたり追跡調査していく研究手法を**コホート研究**と呼びます。

曝露とは「何かを体に浴びる」というようなことではなく、**「特定の状態」**のことです。例えば、喫煙も曝露になりますし、毎朝 30 分の散歩や、野菜を多めに食べるなどの生活習慣も曝露となります。もちろん薬やサプリメントの使用、あるいは年齢や性別なども曝露と考えることができます。つまり、**疾患が発生する以前の「状態」のことを曝露**と呼ぶのです。曝露のなかでも特に疾患の発生に高い確率で影響を与えるものが**危険因子**です。例えば喫煙は肺がんの危険因子と言えるでしょう（図3）。

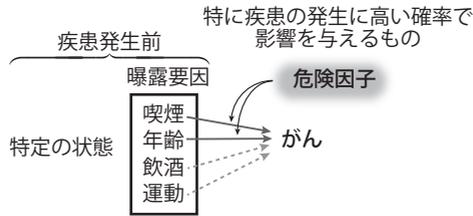


図3 曝露と危険因子

- ・ フラミンガム研究
  - ▶ 米国マサチューセッツ州フラミンガム町の住民
- ・ 米国退役軍人コホート
  - ▶ 米国の退役軍人を対象としたコホート研究
- ・ Medicare Claim Database
  - ▶ 米国の 65 歳以上を対象とした公的医療保険 Medicare 加入者
- ・ General Practice Research Database (GPRD)
  - ▶ 英国の診療所より収集された診療データベース
- ・ 大崎国民健康保険コホート
  - ▶ 宮城県大崎保健所管内の国保加入者
- ・ 多目的コホート研究
  - ▶ 11 カ所保健所管内の住民
- ・ 広島・長崎原爆被ばく者コホート
  - ▶ 広島・長崎の原爆被爆者
- ・ 久山町研究
  - ▶ 福岡県久山町の地域住民

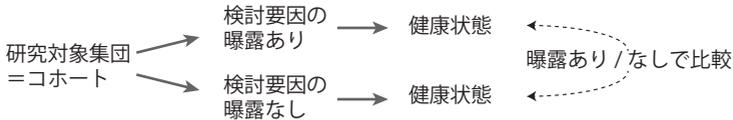
図4 世界各国および日本の代表的なコホート

世界各国で大規模コホートデータベースが構築されており、健康と曝露因子の関連について、様々な研究が行われています(図4)。世界的にも有名なフラミンガム研究は、米国マサチューセッツ州フラミンガム町の住民 5,209 人を対象に 1948 年に開始された大規模コホート研究です。半世紀近くもの追跡調査が行われるなかで、血圧値とコレステロール値が高い患者で循環器疾患が多いことや、喫煙や座りがちの生活なども循環器疾患の発症リスクを増大させる<sup>6-8)</sup>ことが明らかにされ、循環器疾患治療の礎を築きました。

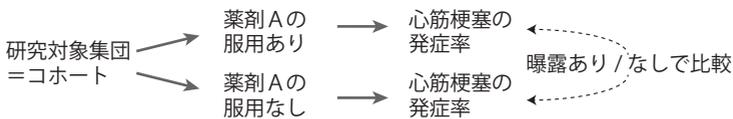
\*3 健康状態の改善もしくは悪化などを検討するための検査値や合併症の発生率、回復率や死亡率など、治療や予防による臨床上の成果をアウトカムと呼びます。

コホート研究では、コホート（研究対象集団）を検討したい曝露（要因や特性）を有している群（曝露群）と有していない群（非曝露群）に分け、健康状態の改善もしくは悪化など<sup>3</sup>を一定期間にわたり観察し、曝露との関連性を検討します。例えば、薬Aを服用することで、どれだけ心筋梗塞の発症リスクが低下するかを検討したい場合、コホートに登録されている集団を、薬Aを服用していた集団と、薬Aを服用していなかった集団の2つに分け、一定期間にわたり追跡調査を行い心筋梗塞の発症率を比較します（図5）。

**【コホート研究のイメージ】**



**【薬剤Aの服用と心筋梗塞発症との関連性を検討する場合】**



**【卵の摂取頻度と心筋梗塞発症との関連性を検討する場合】**



図5 コホート研究の概要

薬Aを服用していた集団を**曝露群**、服用していなかった集団を**非曝露群**と呼びますが、対象となるコホートを曝露群、非曝露群に分け、その2群で健康状態を比較するというのがコホート研究の基本的な構造です。

コホート研究では薬の有効性や副作用リスクのみならず、生活習慣や生活環境なども曝露として、検討することができます。例えば、卵の摂取頻度と心筋梗塞の

関連性を調べたい場合、卵の摂取頻度が低い人（図5では週に1個未満）と頻度が高い人（図5では1日1個以上）を比較して心筋梗塞の発症率を比較することになります。その結果、卵の摂取頻度が高い群で心筋梗塞の発症が多ければ、卵の摂取が心筋梗塞発症に寄与する可能性が示唆されることになります。

コホート研究は大きく、**前向きコホート研究 (prospective cohort study)** と、**後ろ向きコホート研究 (retrospective cohort study)** に分けることができます（図6）。前者は、現時点から未来に向かってデータを収集、解析するのに対して、後者は既存のデータベースを使って（過去の）情報を収集し、現在に向けて解析していきますが、曝露群、非曝露群に分け、その2群で健康状態を比較することに変わりはありません。

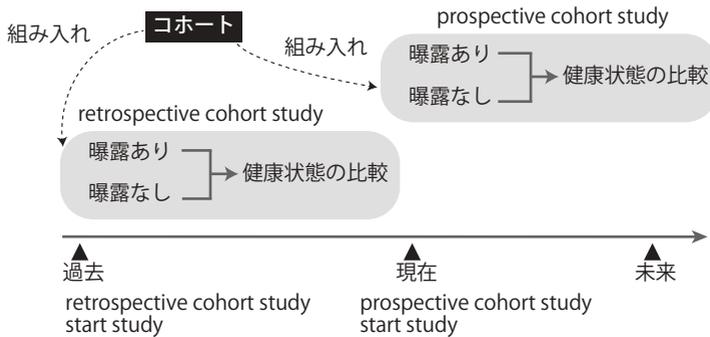


図6 前向きコホート研究と後ろ向きコホート研究

## 2. コホート研究の実例と研究手法の限界

コホート研究は、1人の症例ではなく集団を研究対象にしていること、前後比較ではない比較対照群が設定されている点において、曝露因子の Efficacy に迫ることを可能とした研究手法といえます。しかし、得られた研究結果は必ずしも曝露因子の Efficacy そのものではありません。コホート研究の限界について、実際の研究例を用いてご紹介します。

図7に示したのは、心不全患者に対するレニン-アンギオテンシン系阻害薬(降

圧薬)の影響を検討したコホート研究<sup>9)</sup>です。この研究では、心不全を有する高齢者 2,416 人 (平均 86 歳) を対象に、レニン-アンギオテンシン系阻害薬を服用している 1,208 人 (曝露群) と、そうでない 1,208 人 (非曝露群) が比較され、死亡率が検討されました。

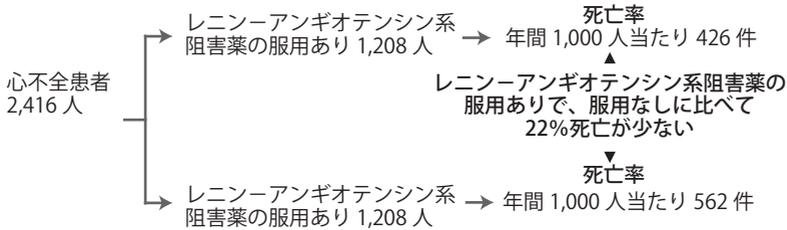


図7 心不全を有する高齢者に対する RAS 阻害薬の影響

(Savarese G, et al: Association between renin-angiotensin system inhibitor use and mortality/morbidity in elderly patients with heart failure with reduced ejection fraction: a prospective propensity score-matched cohort study. Eur Heart J. 2018; 39 :4257-4265)

中央値<sup>4)</sup>で 1.4 年にわたる追跡調査の結果、死亡率はレニン-アンギオテンシン系阻害薬を服用していた曝露群で、年間 1,000 人当たり 426 件、レニン-アンギオテンシン系阻害薬を服用していなかった非曝露群では年間 1,000 人当たり 562 件という結果でした。この発症率の比をとって、薬剤を服用していると 22% 死亡リスクが低下するという結果を導くことができます。すなわち、心不全高齢者ではレニン-アンギオテンシン系阻害薬の服用で、長生きできる可能性が示されているわけです。

レニン-アンギオテンシン系阻害薬の投与前と投与後を比較した研究ではありませんから、「三た論法」が成立している余地はありません。しかし、示された結果は必ずしも薬剤の純粋な Efficacy を示しているわけではありません。例えば曝露群と非曝露群で年齢や喫煙習慣など患者背景に偏りがあればどうでしょう

\* 4 中央値 (median) とは、代表値の一つで、有限個のデータを小さい順に並べたとき中央に位置する値のことです。

か。曝露群で極端に高齢の人が偏っていたら、あるいは喫煙者の割合が高ければ、薬剤の服用とは無関係に曝露群で死亡率やがんの発症率は高くなることでしょう<sup>5</sup>。つまり、曝露群と非曝露群で検討したい曝露以外の背景条件、例えば年齢や性別、生活習慣や運動習慣、社会経済的状況などが、ほぼ均等でなければフェアな比較ができないということです。

図7の研究において、レニン-アンジオテンシン系阻害薬を服用している人と、服用していない人では、生活環境が異なっている可能性が極めて高いように思われます。薬を処方されている集団、つまり医療機関を受診した集団というのは、健康に対する関心が高く、患者本人が身体的・精神的に一定レベルで健康であるか、あるいは周囲の十分な健康サポートが得られる集団だからです。

他方で薬を服用してない非曝露群は、薬を服用する必要のない健康的な人だけではなく、健康に対する関心が低いために医療機関を受診しない人、あるいは生命予後の短い末期状態で、積極的な薬物療法を受けていない人、投与禁忌に該当するような特殊な病状を有する人も含まれています。このような被験者背景の差異が、薬剤の服用とは無関係に、死亡リスクに差をもたらす可能性があります。

この研究では傾向スコアマッチングという解析手法を用いて、曝露群と非曝露群の背景を統計的に補正していますが、一般的なコホート研究において、曝露群と非曝露群の患者特性を完全に均一にさせることは不可能です。結果に影響を与えうる背景因子が既知の因子であれば、あらかじめ補正が可能ですけど、未知の因子では補正しようがありません。コホート研究の結果が、Efficacy そのものを示していない大きな理由は、**曝露群と非曝露群の比較において、被験者背景の差異を完全に補正できない点**にあります。

### 3. 患者背景の差異をどう乗り越えるか?……ランダム化

臨床研究は大きく、**介入研究と観察研究**に分けられます。これまで紹介してき

\*5 背景因子の差異によって結果がゆがめられてしまうことを「交絡が起きている」といいます。詳細は第6のスキルで紹介します。

たコホート研究は、コホートを曝露あり、曝露なしの2群に分け、注目する健康状態に関してその発症率を自然経過の中で「観察」していくという、観察研究の代表的な研究手法です。

他方で、介入研究とは人為的に薬剤を投与するというような、実験的な研究といえます。例えば薬剤効果を検討する場合においては、「薬を飲ませる群」、「薬を飲ませない群（あるいはプラセボを投与する群）」の2つを比較します。コホート研究のような観察研究と大きく異なるのが、**薬を服用することが「曝露」ではなく「医学的介入」となっている点**です。つまり、薬を飲ませる群が**介入群**であり、薬を飲ませない群（プラセボ投与群）が**非介入群**となります<sup>6</sup>。

そして、介入研究の大きなメリットの一つは、**ランダム化**という手法を用いることで、介入群と非介入群における患者背景の偏りを防ぐことができることです。コホート研究では曝露群、非曝露群への振り分けは、患者自身の生活環境に存するもので、研究者の意思は介在しません。しかし介入研究では、被験者を研究者の意思で介入群と非介入群へ割り当てることができます。そして、被験者を意図的にランダム（でたらめに）に2つの群へ振り分けることで、介入以外の背景因子、年齢や性別、生活習慣や社会経済的環境などを均一にすることが可能となります<sup>7</sup>。

例えば、サイコロを振って、偶数が出れば介入群へ、奇数が出れば非介入群へ割り付ければ、一定数の集団を対象とした場合においては、被験者の背景因子をほぼ均等に2群に割り付けることができるということです。このように、被験者集団を介入群、非介入群の振り分ける際、人工的に偶然を発生させ、ランダムに割り付けていく手法を**ランダム化**と呼びます<sup>8</sup>。

---

\* 6 「薬の服用」を介入とするか曝露とするかで研究デザインはまるで異なったものになりますし、研究結果の内的・外的妥当性も大きく変化します。

\* 7 この場合、統計的補正と異なり、結果に影響し得る未知の因子についても、介入群と非介入群で均等に振り分けることが可能です。

\* 8 ランダム化を無作為化と呼ぶことがありますが、ランダムに割り付ける過程において、既に作為が存在すると筆者は考えています。つまり意図的にランダム化しているということです。研究者の意図が介在しなければ、ランダム化を行うことはできず、それはまた、自然選択的な状況を維持するという意味において観察研究との差異を見いだせません。

ランダム化を行うことで、研究者側も、被験者が介入群、非介入群どちらの群になるか予想することが困難となります。このことはまた Efficacy を検出するうえでの利点となります。例えばランダム化を行わず、研究者が非介入群に不健康な人たちを設定し、介入群に健康的な人たちを設定してしまうと、医学的介入の効果とは関係なく、非介入群での死亡率は高くなってしまいうでしょう。このように、ランダム化という手法を用いて2群を設定することにより、比較したい治療以外の背景因子を、2群間で均等に振り分けることが可能となります (図8)。

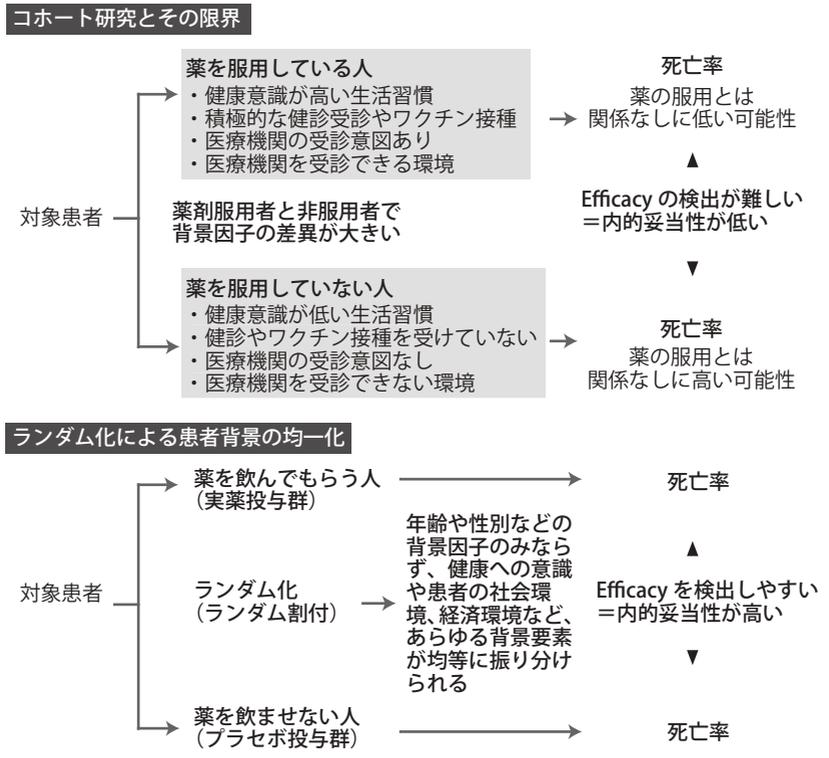


図8 コホート研究の限界とランダム化比較試験

## 4. ランダム化比較試験とその限界

被験者集団を介入群、非介入群にランダム化し、一定期間追跡調査を行うことで、介入効果を検討する研究手法を**ランダム化比較試験 (Randomized controlled trial : RCT)**と呼びます。ランダム化比較試験は、集団を対象にしていること、介入以外の背景因子が介入群とほぼ同質な比較対照群を設定していることから、Efficacy 検出に優れた研究デザインです。したがって、一般的に医学的介入効果の検証を行う際にはランダム化比較試験が実施されます。

とはいえ、ランダム化比較試験にも研究デザイン上の限界は存在します。図9に示したのは、心臓病のリスクが高い2型糖尿病患者を対象に、エンパグリフロジン（血糖降下薬）の心臓病に対する有効性を検討したランダム化比較試験<sup>10)</sup>の概要です。研究の結果、心臓病の発症は、エンパグリフロジンを投与した群で10.5%、プラセボ群では12.1%でした。その比をとると0.86となり、エンパグリフロジンはプラセボに比べて14%心臓病のリスクを減らせることが示されています。

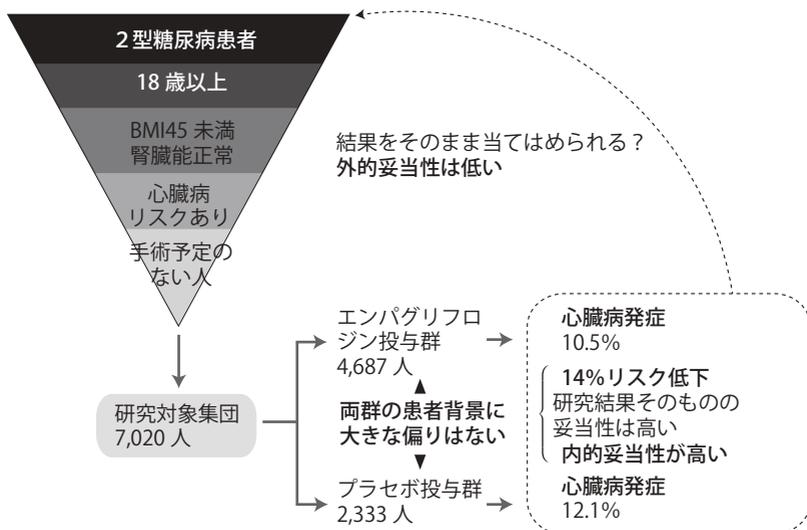


図9 エンパグリフロジンの有効性・安全性を検討したランダム化比較試験 (Zinman B, et al: Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med. 2015; 373: 2117-2128)

この研究は7,020人という集団を対象としていること、さらには患者背景に偏りのない2群を比較していることから、示された結果はエンパグリフロジンのEfficacyに近しいものだと考えることができます。しかし、この結果はあくまで研究対象集団でのデータであることに注意が必要です。この結果が2型糖尿病患者すべてに妥当するかどうかについては議論の余地があります。

介入研究とはある意味で実験的な介入でした。したがって、健康状態が著しく悪い集団や、薬の使用が禁忌に該当する集団、重篤な合併症を有する集団は倫理的な観点から研究に組み入れません。実はランダム化比較試験には厳格な研究組み入れ基準が設定されており、被験者の安全性や倫理性に配慮されているのです。また、最終的に同意を得られた人でしか研究が行われませんので、被験者集団は一般的な集団とはやや異質な集団ということができます。

## 5. 内的妥当性と外的妥当性

ランダム化比較試験の限界について理解するうえで、**内的妥当性と外的妥当性**を理解しておく必要があります。繰り返しますが、ランダム化比較試験で示された研究結果は、介入のEfficacy（因果効果）に近しいものでした。つまり、介入という「原因」がもたらす健康状態への影響という「結果」の連関、因果関係の強さが定量的に示されているわけです。

このように、示された研究結果と介入（もしくは曝露）が因果関係にあるかどうかということに関する妥当性を、**内的妥当性**と呼びます。内的妥当性は**研究対象集団における研究結果の再現性の高さ**と考えればわかりやすいかと思います<sup>9</sup>。そして、ランダム化比較試験は観察研究と比較して内的妥当性に優れた研究デザインです。

内的妥当性に対して、示された研究結果を広く一般化できるかどうかに関する度合いを**外的妥当性**と呼びます。つまり、研究結果における普遍性の高さです。

---

\*9 研究対象者と全く同じ集団に対して、全く同様の介入を行った場合、同等の結果が再現される程度のこと、と言い換えてもよいでしょう。

人を対象とした臨床研究は一般的に**標本調査**です。研究対象に該当するすべての人、つまり母集団すべてを集めて研究する**全数調査（悉皆調査）**ではなく、研究参加に同意し、研究の組み入れ基準を満たした特定の集団を対象（これを標本と呼びます）に研究を行います。得られる結果の妥当性という観点からいえば母集団すべてを対象とした全数調査が望ましいですが、例えば糖尿病の研究をするにあたり、地球上に存在するすべての糖尿病患者を対象に研究することなど不可能でしょう。

数万人規模の症例を解析できる観察研究では、標本は母集団の特性と近似でき、結果の外的妥当性、つまり得られた研究結果の一般化可能性は高いと考えられます。他方、ランダム化比較試験では、研究の組み入れ基準が厳格であること、またコストの面でも大規模な研究は困難であることが多く、標本数は多くても数千から1万人規模です。したがって、標本集団は母集団との背景特性がかけ離れていることも多く、結果の外的妥当性は観察研究と比較すると低下します（図10）。

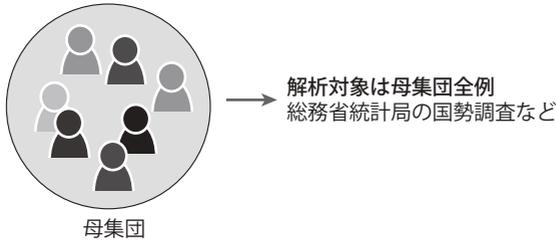
実際、2型糖尿病患者を対象としたグルカゴン様ペプチド（GLP）-1 受容体作動薬の主要なランダム化比較試験において、米国における一般的な2型糖尿病患者の多くが研究に組み入れられていないことが報告されています<sup>11)</sup>。各ランダム化比較試験における、一般的な2型糖尿病患者組み入れ割合の推定値は、EXSCEL 試験で15.9%、SUSTAIN-6 試験で13.0%、LEADER 試験で12.9%という結果で、実際には心血管疾患ハイリスクな患者が多く組み入れられていました（表6）。

表6 研究に組み入れられた米国2型糖尿病成人の割合推定

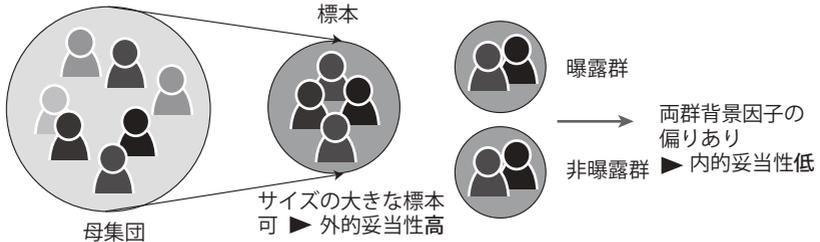
	成人2型糖尿病	EXSCEL	LEADER	REWIND	SUSTAIN-6
症例数	26110573	14752	9340	9901	3297
平均年齢	60.5 ± 13.5	62.7 ± 9.9	64.3 ± 7.2	66.2 ± 6.5	64.6 ± 7.4
男性	55.5%	62.0%	64.3%	53.7%	60.7%
BMI	33.2 ± 6.7	31.8 ± 5.9	32.5 ± 6.3	32.3 ± 5.7	32.8 ± 6.2
eGFR < 60	17.3%	18.6%	<b>23.1%</b>	<b>22.2%</b>	<b>28.5%</b>
HbA1c	7.2 ± 1.6	<b>8.0 ± 1.2</b>	<b>8.7 ± 1.5</b>	7.3 ± 1.1	<b>8.7 ± 1.5</b>
心筋梗塞既往	5.5%	—	30.7%	16.2%	32.5%
一般的な成人2型糖尿病患者の割合		<b>15.9%</b>	<b>12.9%</b>	42.6%	<b>13.0%</b>

(Imfeld P, et al: Benzodiazepine use and risk of developing Alzheimer's disease or vascular dementia: a case-control analysis. Drug Saf. 2015; 38: 909-919)

【全数調査（悉皆調査）】



【コホート研究（標本調査）】



【RCT（標本調査）】

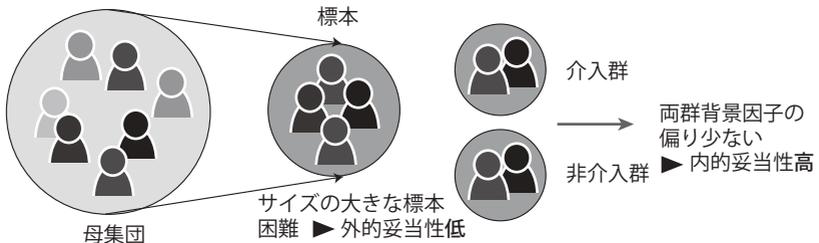


図 10 研究の外的妥当性

## 6. 症例対照研究

Efficacy を検出するための研究手法で重要なのは、**個人ではなく集団を研究対象**としていること、**適切な比較対照群が設定されていること**の2点であり、この条件を満たす代表的な研究手法がコホート研究、そしてランダム化比較試験でした。

この2つの研究手法に共通するのは、一定数の被験者集団を、介入あり／なし、