

はじめに

ChatGPTに代表されるような生成AI (Generative Artificial Intelligence: 生成型人工知能) の登場によって、いわゆるホワイトカラー職種の多くが、一時解雇 (レイオフ) の対象となっている現実があります。たとえば、米国のマイクロソフト社は、AIに合わせた事業モデルへの転換を進めるなかで、約9000人規模の人員削減を発表しています(日本経済新聞、2025年7月3日付)。

かつて、インターネットの登場と普及が、情報の扱い方や私たちの生活様式を根本から変えたように、生成AIの普及もまた、あらゆる分野に影響を及ぼしつつあります。

医療者の業務環境もまた、生成AI技術の発展にともない、自らの専門性と職業的な価値を根本から見直すべき時を迎えているように思います。とりわけ、医薬品に関する情報を収集・活用するDI (Drug Information) 業務は、生成AIの普及によって大きく様変わりしていくことでしょう。英語で書かれた医学論文を読むことのハードルは大きく低下し、エビデンスと呼ばれる論文情報がさまざまな文脈で、当たり前にも語られる日はそう遠くないように感じています。

しかしながら、情報の収集や整理と、情報から意味をみだし、意思決定に活かすことは全く異なる営みです。前者はインフォメーションの取り扱いですが、後者はインテリジェンスと呼ばれるものでしょう。そして、このインテリジェンスを生み出すために不可欠なスキルが**情報分析力**です。

生成AI時代において医療者に求められることの一つは、医薬品やヘルスケアに関する論文情報を批判的に読み解き、文脈に即した意味を創造する力と倫理的な配慮です。その意味では、DI業務はDrug Informationではなく、Drug Intelligenceと呼ぶべきなのかもしれません。

本書では、臨床医学に関する論文を批判的に読み解くために必要なスキルを、問題と解説の形式で網羅的に紹介しました。さらに、各問題に

対する回答を生成AIツールである ChatGPT に作成させ、著者による解説と対比することで、情報分析力を実践的かつ効果的に養える構成としました。

生成AIの登場は、医療者にとって脅威であると同時に、職業的専門性を再構築し、より高度な価値を創出する機会でもあります。定型的かつ機械的な業務は、技術の進化により確実に自動化されていく一方、人間にしか担えない判断や価値創造の領域は、相対的に重要性を増していくでしょう。本書で紹介した情報を批判的に読み解くスキルが、Drug Intelligenceを実現するための基盤となり、臨床現場における確かな判断や行動を支える一助となりましたら幸いです。

2025年8月 青島周一

目次

はじめに	2
本書の構成	7
本書を読むうえで整理しておきたい基礎知識	10
序 医学論文を読む理由と本書のコンセプト	18

I

ランダム化比較試験論文の ピットフォール(基礎編)

1 臨床医学研究における比較方法の重要性	24
2 盲検化が困難なRCTとPROBE法	31
3 代用のアウトカムと真のアウトカム	39
4 仮説検証と仮説生成の違い	46
5 複合エンドポイントと、その解釈の視点	52
6 統計学的に有意な差がないとはどういうことか	59
7 ランダム化比較試験の外的妥当性	65
8 二重盲検が適切に機能していない可能性を探る	73
9 NNTの考え方と解釈のピットフォール	78

II

ランダム化比較試験論文の ピットフォール(実践編)

10	臨床試験で実施された治療介入の再現性	86
11	研究の早期中止と解析結果の過大評価	95
12	標準治療の異質性がもたらす治療効果の減弱	102
13	複合エンドポイントとハザード比のピットフォール	108
14	ランダム化比較試験における効果量と倫理的問題	114
15	治療に対する期待感と薬剤効果の多因子性	120
16	被験者の組み入れ方法と薬効感に対する影響	127
17	臨床研究における社会的望ましきバイアス	133

III

観察研究論文のピットフォール

18	コホート研究における交絡バイアスと選択バイアス	142
19	コホート研究における初期症状バイアス	151
20	コホート研究におけるImmortal time bias	160

21	薬物有害事象の評価と発見兆候バイアス	171
22	有名医学誌の掲載論文でも鵜呑みにできない理由	178
23	医療データベースを用いた研究のピットフォール	188

IV

システマティックレビュー・ メタ分析論文のピットフォール

24	メタ分析の結果と加重平均ウェイト	196
25	メタ分析による有害事象評価の留意点	207
終	AIが語る“それっぽさ”にどう抗うか	216

本書の構成

臨床研究の概要と、批判的吟味に必要なノウハウに関する問題

問題に回答するにあたって参照すべき統計データ

取りあげた論文のPubMed抄録へアクセス

1 臨床医学研究における比較方法の重要性

問題 桔梗湯^{ききょうとう}は、扁桃炎や扁桃周囲炎に適用を有する漢方薬で、咽頭痛の緩和を期待して処方されることが一般的である。

2013年に桔梗湯の有効性を検討した臨床試験の結果が報告されている。この研究では、咽頭痛を訴える40人が対象となった。被験者の咽頭痛症状について、100点満点のスコア（点数が高いほど症状が重い）で評価した後に、桔梗湯2.5gをコップ1杯のお湯で投与した。投与の10分後に咽頭痛症状を再評価し、桔梗湯の投与前後におけるスコア変化が比較されている。

その結果、スコアは投与前後で統計学的にも有意に減少した。この研究結果を根拠に、桔梗湯が咽頭痛の症状を改善するといえるだろうか。

●論文データ

- 出典：J Complement Integr Med. 2013 Dec 20; 11 (1): 51-4.
- PMID：24356393
- 研究デザイン：前後比較試験
- 解析対象：急性上気道炎による咽頭痛を訴える日本人40人（平均34.6歳）
- 主な結果：疼痛症状の平均スコアは投与前48.2点と、投与10分後35.4点と、投与前後で統計的にも有意に減少（ $p < 0.001$ ）

PubMed抄録へアクセス→



24

問題に対する著者による解答

解答

この研究結果は、桔梗湯の薬理学的な作用以外の要因、すなわちプラセボ効果や自然治癒などの影響を多く含んでいる。症状スコアの低下は桔梗湯そのものの効果とはいいがたく、この結果を根拠に咽頭痛の症状を改善するとはいえない。のど飴やハチミツでも、同様の結果が得られた可能性を指摘できる。



・・・ ChatGPTの回答例

桔梗湯が咽頭痛の症状を改善するといえるかどうかについて、この研究結果を批判的に吟味するためには、研究デザイン、バイアスの可能性、内的および外的妥当性、そして結果の臨床的

I

ランダム化比較試験論文のビッドフォール（基礎編）

解答

この研究結果は、桔梗湯の薬理学的な作用以外の要因、すなわちプラセボ効果や自然治癒などの影響を多く含んでいる。症状スコアの低下は桔梗湯そのものの効果とはいいがたく、この結果を根拠に咽頭痛の症状を改善するとはいえない。のど飴やハチミツでも、同様の結果が得られた可能性を指摘できる。



ChatGPTの回答例

桔梗湯が咽頭痛の症状を改善するといえるかどうかについて、この研究結果を批判的に吟味するためには、研究デザイン、バイアスの可能性、内적および外적妥当性、そして結果の臨床的意義を考慮する必要があります。

まず、本研究は**前後比較試験**というデザインで実施されており、これは被験者自身をコントロールする方法です①。このデザインの利点として、個人間のバリエーションを排除できる**点が挙げられますが、プラセボ効果や自然経過の影響を排除することが難しいという限界**もあります。とくに咽頭痛のような症状は自然軽快することが多く、さらに温かい飲み物そのものが鎮痛効果をもつ可能性も考慮する必要があります②。この研究では対照群が設けられておらず、プラセボ効果や温熱効果を適切に評価することができていません。

次に、内적妥当性について考えると、評価指標として視覚的アナログスケール（VAS）が用いられ、投与前後で有意なスコアの減少が示されています（ $p < 0.001$ ）。しかし、VASは主観的な評価方法であり、被験者が介入によって症状が改善すると期待することでスコアが変動する可能性があります。また、測定時点が投与後10分および30分と短時間であることから、持続的な効果や再発に関する情報は得られていません。

外적妥当性については、研究参加者が単一施設から募集されており、対象者の年齢や基礎疾患、症状の程度に偏りがある可能性が指摘できます。このことは、結果の一般化可能性を制限します。また、サンプルサイズが40人と小規模であるため、**偶然の影響や統計的誤差の可能性も否定できません** [Point ①]。

さらに、臨床的意義についても考慮する必要があります。統計的には有意な差が示されていますが、平均VASスコアの変化は約13ポイントの減少に留まっており、これが**実際に患者にとって意味のある改善であるかは不明**です [Point ②]。臨床的に有意な変化（Minimal clinically important difference: MCID）がどの程度かについては議論の余地があり、この研究ではその基準が示されていません。

以上の点から、この研究結果だけを根拠に桔梗湯が咽頭痛の症状を改善すると断定するのは難しいといえます。**プラセボ対照無作為化試験**など、より厳密な研究デザインによる追加のエビデンスが求められます③。

※生成AIの回答には不正確な内容が含まれている場合もあります。

※回答生成は2025年2月3日

ChatGPTによる回答

強調を目的としてAI回答の一部を太字に加工しています

後出の著者による解説に対応する番号です

記載内容の注意点

Point

Caution!

ChatGPTによって出力された文章は、基本的には原文のまま掲載していますが、上記のほか、専門用語の統一（たとえば、アウトカムをエンドポイントに統一）や「ですます調」への変換など、若干の修正を行っています

著者による解説

丸数字や注意点、注目点はCharGPTの回答部分と対応しています

本書に掲載されている図表のうち、出典がないものは全て著者作成です

●解説

1. 前後比較試験の問題点

薬や検査など、医学的な手段（介入）の効果は、その手段を行った場合と行わなかった場合の差異で示すことができます。問題で紹介した研究は、**桔梗湯を服用する前**（手段を行わなかった場合）と**服用した後**（手段を行った場合）を比較することで、その効果を検討しています。このような研究方法を**前後比較試験**と呼びます①【図1-1】。

前後比較試験では、介入の前後を比較するだけなので、研究方法としては簡便です。しかし、得られた結果の信頼性（内的妥当性）は低く、治療効果の検証に用いるべきではありません。その理由は、「てるてる坊主とその効果」を想像するとわかりやすいと思います。てるてる坊主を軒先に吊るしたところ、降り続いていた雨が翌朝に止んだとしましょう。この場合、てるてる坊主が晴天をもたらしたと結論できるでしょうか。常識的な直観によれば、「てるてる坊主に晴天をもたらす効果がある」とは考えにくいと思います。気象現象は気温や湿度、地域性などさまざまな要因によって生み出される**複雑な因果連鎖の流れ**です。そういう意味では、人の健康状態もまた、年齢、生活習慣、居住地域、医療に対する関心など、複雑な因果連鎖の流れに支配されているといってもよいでしょう。

「薬を服用した」「症状が改善した」、だから「薬には効果があった」という三段論法は、「てるてる坊主をつるした」「翌朝晴れた」、だから「てるてる坊主には効果があった」という三段論法と同じ構造です。この三段論法で無視（あるいは軽視）されているのが、**薬（てるてる坊主）以外の要因**です。



図1-1 前後比較試験のイメージ

I

ランダム化比較試験論文のピットフォール（基礎編）

1

臨床医学研究における比較方法の重要性

27

Minimal clinically important difference (MCID) とは、臨床的に意味のある最小の変化差を表す概念です。本研究では、MCIDが事前に設定されておらず、ChatGPTの指摘通り、「平均VASスコアの変化は約13ポイントの減少に留まっており、これが実際に患者にとって意味のある改善であるかは不明」です【Point②】。

【参考文献】

1) Intern Med. 2019 Sep 1; 58 (17): 2459-2465. PMID: 31178508

参考文献

PMIDやDOIのナンバーから論文を検索できます

序

医学論文を読む理由と 本書のコンセプト

臨床現場で医療者が抱く疑問を**臨床疑問 (Clinical question)**と呼びます。研修中の医学生を対象とした研究によれば、**患者1人当たり5つの臨床疑問が提起**されると報告されています¹⁾。薬剤師の臨床現場においても、薬物療法の有効性や安全性、食事療法や運動療法の効果など、さまざまな疑問を抱くことでしょう。

医学論文を読まなくてはならない理由

臨床疑問は、**背景疑問 (Background question)**と**前景疑問 (Foreground question)**に分けられます【図0-1】²⁾。

背景疑問とは、自分が知らないだけですでにわかっている**一般的な知識に関する疑問**のことです。自分が知らない分野について学ぼうとする際に、初めて遭遇する疑問といってもよいかもしれません。

たとえば、**薬の作用機序**や**疾病の病態生理**に関する疑問は典型的な背景疑問です。あるいは、**専門用語の定義に関する疑問**は、最もシンプルな背景疑問といえるでしょう。このような疑問の多くは、標準的な教科

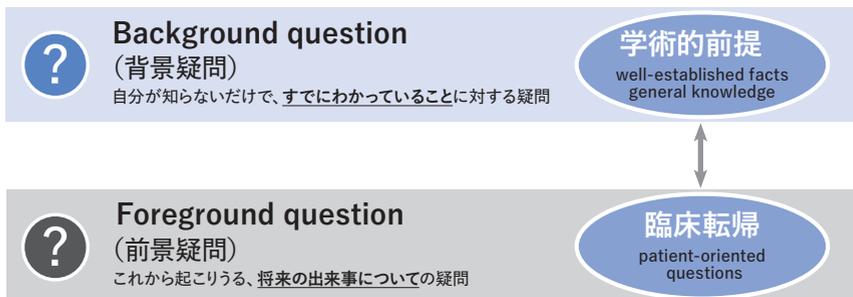


図0-1 背景疑問と前景疑問

書などを参照することで解決できます。

他方の前景疑問とは、**将来的に起こりうる患者個別の出来事に関する疑問**です。たとえば、ある薬を目の前の患者に投与すると、どれだけ心血管疾患の発症を予防できるのか、という疑問を挙げることができます。これから先、**未来に起こる出来事についての疑問**であるがゆえに、前景疑問に対する明確な答えは、問いを立てた時点では存在しません。

前景疑問は、将来の健康状態に関する疑問であり、将来の出来事を合理的かつ客観的に見積もるためには、**統計学的なデータが必要不可欠**です。このようなデータは、人を対象とした臨床研究によって得られることが一般的であり、その結果は**医学論文**として報告されます。

なお、医療・保健衛生分野でエビデンスといった場合、臨床研究の結果を報告した医学論文（以下、医学論文とします）を指します。そして、医学論文に記載されている統計学的なデータを踏まえながら、臨床現場の状況や環境、患者の意向や価値観、医療者の臨床経験を統合した意思決定アプローチを**EBM (Evidence-based medicine)**と呼びます^{2) 3)}。

EBMは5つのステップから構成される医療者の臨床行動スタイルです【図0-2】²⁾。ステップ1では、前景疑問を「**どんな患者に、どんな治療をすると、何と比べて、どうなるか**」という形式で定式化します。なお、どんな患者に (Patients)、どんな治療をすると (Intervention)、何と比べ

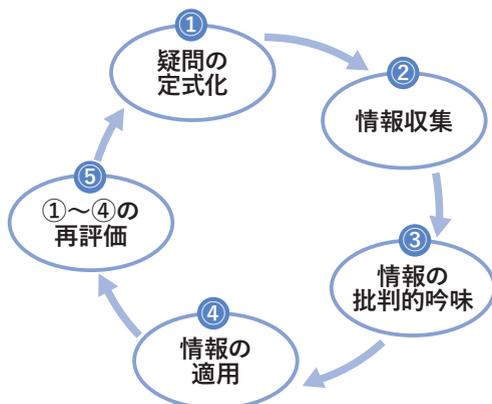


図0-2 EBMの5つのステップ

て (Comparison)、どうなるか (Outcome) は、前景疑問を構成する4つの要素です。それぞれの英単語の頭文字をとってPICOと呼びます。

ステップ2では、疑問の解消に参考となるような情報を検索します。優先的に参照すべき情報は、多くの場合で**医学論文**、もしくは**医学論文を引用した専門資料**です。

ただし、医学論文に記載された統計学的なデータには、偶然やバイアスの影響が少なからず含まれています。そのため、医学論文の内容は、統計学的なデータも含めて**批判的に吟味 (Critical appraisal)** する必要があります。この吟味プロセスこそがEBMのステップ3に該当します。

医学論文等の批判的な吟味によって得られた示唆は、EBMのステップ4において、患者の意向や価値観、臨床現場の状況や環境、医療者の経験という3つの要素を加味して、最終的な臨床判断に活用されます。なお、医学論文の結果とこれら3つの要素を、**臨床判断の4要素**と呼びます。一連の行動プロセスに対して、修正すべき点や付け加えるべき点などを評価し、改めて前景疑問を提起するのがEBMのステップ5です。

本書のコンセプト

本書は、EBMのステップ3に該当する**医学論文の批判的吟味**を、効率的かつ網羅的に学習できるように、押さえておくべき重要論点を**問題形式**でまとめています。各章の冒頭には、医学論文で報告されている臨床研究の概要と、批判的吟味に必要なノウハウに関する問題が提示されています。また、問題を回答するにあたって参照すべき統計データは、【論文データ】の項目にまとめています。そのため、実際の医学論文を閲覧しなくても、本書のみで批判的吟味のスキルが習得できるよう配慮しています。

ただし、【問題】で取り上げた医学論文の多くは、インターネット上で無料閲覧できます。【出典】に記載したPMIDを論文検索サイトである**PubMed**に入力するか、【問題】下部に記した**二次元バーコード**を読み込むことで、該当論文の抄録にアクセスできます。PubMedとは、世界中の医学文献データベースであるMEDLINE (メドライン) をインター

ネットで検索できる無料のデータベースで、<https://pubmed.ncbi.nlm.nih.gov/>からアクセスできます。必要に応じて、実際の論文を読みながら【問題】に対する回答を考えてみると、より理解が深まるでしょう。

とはいえ、論文を読みなれていない方や、英語が苦手な方にとって、医学論文の批判的吟味は大変な作業かもしれません。近年では、ChatGPTのような生成AIツールが無料で使えるようになりましたので、論文のPDFファイルを生成AIに分析させてみることで、批判的吟味のハードルを大きく下げられるように思います。

その際には、①研究の背景と研究の目的を解説してください、②研究の概要をPICOの形式で出力してください、③主な研究結果を整理したうえで、その内的妥当性を批判的に吟味してください、④EBMのステップ4における臨床判断の4要素を踏まえて、研究結果の適用を考察してください、などのプロンプト（生成AIに対する指示）を使うことで、効率的に医学論文を読むことができるはずです。

本書の各章で出題した【問題】の答えについても、生成AI（ChatGPT）による回答と、著者による解答を併記しています。生成AIによる回答は、【問題】で取り上げている医学論文のPDFファイルをChatGPTにアップロードし、以下のプロンプトを用いて出力させました。

以下の【問題】は、医療従事者を対象に、添付した医学論文の批判的吟味（EBMのステップ3）を学ぶ目的で作成されている。添付した学術論文を体系的に分析したうえで、【問題】に対する回答を作成せよ。なお、回答にあたっては、箇条書きを可能な限り使用せず、流暢な日本語でわかりやすい文章にせよ。

【問題】

各章の記載内容をコピー＆ペースト

【論文データ】

各章の記載内容をコピー＆ペースト

ChatGPTによる批判的吟味と、著者による批判的吟味を比較するこ

とで、**生成AIを使った医学論文の読解ノウハウや、医学論文の批判的吟味を生成AIで行ううえでの留意事項**を把握できるでしょう。

むろん、ChatGPTをはじめとした生成AI技術の進化は目覚ましく、プロンプトに対する回答内容の質は、月単位で向上している可能性も指摘できます。生成AI技術の進化と、回答内容の質の変化については、本書の終章で考察します。

【参考文献】

- 1) Ann Intern Med. 1991; 114 (7): 576-81. PMID: 2001091
- 2) Straus SE, Glasziou P, Richardson WS, et al. Evidence-Based Medicine: How to Practice and Teach It (4th ed.) . London: Churchill Livingstone; 2010.
- 3) BMJ. 2002; 324 (7350): 1350. PMID: 12052789

著者紹介

青島 周一（あおしま しゅういち）

城西大学薬学部卒業後、保険薬局勤務を経て2012年より医療法人社団徳仁会中野病院勤務、特定非営利活動法人アヘッドマップ共同代表。

公式ウェブサイト (<https://syuichiao.wixsite.com/website>)

著書に『薬の現象学 存在・認識・情動・生活をめぐる薬学との接点』（丸善出版、2022年）、『エビデンスをめぐる往復書簡—EBM実践の向こう側』（共著、中外医学社、2022年）、『OTC医薬品どんなふう販売したらイイですか？ 「全くない」と「ほとんどない」の間にある、ふわふわした効果を探す物語』（金芳堂、2021年）など。

主な執筆論文：

- J Gen Fam Med. 2017 Jun 21; 18(6): 393-397. PMID: 29264070
- J Gen Fam Med. 2021 May 28; 22(4): 182-196. PMID: 34221792
- Pharmacy (Basel). 2024 Dec 5; 12(6): 184. PMID: 39728849

生成AIと学ぶ医学論文の読み方・考え方

2025年9月25日 第1刷発行

著 者：青島周一

発 行：株式会社薬事日報社

〒101-8648 東京都千代田区神田和泉町1番地

電話 03-3862-2141 FAX 03-3866-8495

ホームページ <https://www.yakuji.co.jp>

オンラインショップ <https://yakuji-shop.jp>

組版・デザイン：株式会社ダイヤモンド・グラフィック社

印刷・製本：株式会社ダイヤモンド・グラフィック社

© Syuichi Aoshima, 2025

Printed in Japan. ISBN978-4-8408-1659-5

落丁本、乱丁本はお取り替えます。

 <(社)出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構（電話03-5244-5088、FAX 03-5244-5089、e-mail: info@jcopy.or.jp）の許諾を得てください。