

厚生労働科学研究費補助金（臨床研究基盤整備推進研究事業）

平成20年度 研究協力報告書

臨床研究フェローシップ構築に関する研究

「臨床研究抄録ブラッシュアップセミナー」ワークショップ 報告

研究協力者	横山 葉子	（京都大学大学院医学研究科 医療疫学）
	竹上 未紗	（京都大学大学院医学研究科 医療疫学 助教）
	小崎真規子	（北野病院、京都大学医学研究科 医療疫学 非常勤講師）
	杉岡 隆	（京都大学大学院医学研究科 医療疫学）
	川口 武彦	（京都大学大学院医学研究科 医療疫学）
	島田 利彦	（京都大学大学院医学研究科 医療疫学）
	山本 洋介	（京都大学大学院医学研究科 医療疫学）
	佐久嶋 研	（京都大学大学院医学研究科 医療疫学）
	宮下 淳	（音羽病院、京都大学大学院医学研究科 医療疫学）
	岡林 里枝	（京都大学保健管理センター、奈良県立医科大学総合医療学）
	岡村真太郎	（天理よろづ病院、京都大学医学研究科 医療疫学）
主任研究者	福原 俊一	（京都大学大学院医学研究科 医療疫学 教授）

研究要旨

医師を対象に、リサーチ・クエスションの構造化・英文抄録の作成方法の本質・基本事項の理解を促進する目的で、ワークショップを行った。本ワークショップは、学習効果の向上を目的に、実際の学会抄録を用いるという新規教育モデルを企画し、実施した。参加者からは、高い評価・満足度が得られ、抄録採択者からも良好なフィードバックを得ることができた。本ワークショップは、今後臨床研究のスキルを向上するための、効果的な教育モデルとして有効であることを示唆した。

A. 研究目的

実際の抄録を題材に、前半にリサーチ・クエスションの構造化、後半に英文抄録作成を行ない、英文抄録を臨床研究デザイン・英語表現の両者の点からブラッシュアップすることを目的とした。

B. 研究方法

日時：2008年6月8日（日）10時から16時までの4時間のワークショップを京都大

学大学院医学研究科G棟セミナー室Aで開催した。

- ・ワークショップ開催前に各自英文抄録を提出させた。
- ・前半・後半にそれぞれリサーチ・クエスションの構造化、英文抄録の作成方法に関するミニ・レクチャーを行なった。
- ・提出された英文抄録の中から採択した抄録をもとに、前半はリサーチ・クエスションの構造化をグループワークで行

った。後半も同じ題材を用いて英文抄録作成を個人ワークで行った。

ワークショップのプログラムと内容の詳細を以下に示す。

(1) ミニレクチャー (35 分) (講師: 福原俊一)

ミニレクチャーでは、臨床の漠然とした疑問をリサーチ・クエスチョンに構造化し、説得力のある抄録をつくることを目的とし、以下の項目についてレクチャーを行った。

- ・ 臨床の漠然とした疑問から、リサーチ・クエスチョンに構造化し、構造化抄録を作成するまでの 7 つのステップ
- ・ 診療上の疑問を RQ に構造化する
- ・ RQ をもとに概念モデルを作成する

(2) グループワーク説明 (10 分) (担当: 小崎真規子)

(3) グループワーク (50 分)

(ファシリテーター: 竹上未紗、小崎真規子、杉岡隆、川口武彦、島田利彦、山本洋介、佐久嶋研、宮下淳、岡林里枝)

参加者を 4-5 人のグループ 8 つに分け、各グループ、ファシリテーター 1 名を加え、以下についてグループワークを行った。

- ・ 自己紹介、役割分担 (5 分)
- ・ 現在の抄録の改善点 (10 分)
- ・ 改善点をふまえ、RQ を絞る (10 分)
- ・ 質疑応答 (15 分)
- ・ 改善点を反映した PECO (15 分)
- ・ 発表準備 (5 分)

なお、グループワークの途中で抄録採択者に対する質疑応答の時間を設け、抄録採択者の臨床の疑問が何であったのかを共有し、研究の実施可能性に際して理想的なデザインが行えなかった理由、などを聞く時間と

した。

(4) グループ発表と質疑応答 (20 分)

グループディスカッションの内容の発表を行った。

演者とグループとのディスカッションも適宜行った。

(5) 総評 (20 分) (講師: 福原俊一、杉岡隆)

グループディスカッション、発表を踏まえ、今回の採択抄録を選択した理由、リサーチ・クエスチョンを構造化する際のポイントについて、総評を行った。

抄録を外側から磨く: 英語表現編

(6) ミニレクチャー (30 分): ブライアン・バジエル

- ・ 英文抄録の書き方

(7) 個人ワーク

- ・ タイトルと目的を英語になおす (10 分)
- ・ 背景・方法・結果を英語になおす (20 分)

(8) 発表、コメント、クイズ (75 分): ブライアン・バジエル

- ・ 個人ワークの結果に対するコメント、ミニ・レクチャーの達成度を確認するクイズを行なった。

(倫理面への配慮)

抄録を採択する際には、発表者に事前に連絡をし、ワークショップでの使用可否、名前の提示有無、について事前に確認した。

事後に行ったアンケート調査では、個人が特定されない形で発表されることを明記した。

C. 研究結果

参加者は38名であった。専門は内科系が30名、外科系が5名、不明3名である。

実際の抄録を用い、採択した抄録の演者もワークショップに参加した。

また、グループワークの中に、演者への質疑応答の時間を設けることにより、演者の考えをフィードバックすることができた。

参加者に対するアンケート調査でも、高い評価・満足度が得られた。アンケート調査結果を資料に示す。

さらに、採択抄録の演者からも、「専門科が異なる人からの意見は非常に新鮮であり、新たな気づきがあったように感じます。」といったコメントが得られた。

対象者からのコメントも資料に示す。

D. 考察

初年度は、広範な対象のAwarenessを高めるため、多人数のセミナーを開催した。2年は、目的と対象を絞り込み、より少人数のワークショップ形式を用いた。また、実際の学会抄録を用いるという新規企画は、臨場感のある議論を促進することにつながり、学習効果を向上させると考えられた。このため、三年次には、実際の学会抄録をリサーチ・クエスションの構造化、英語表現という双方からのブラッシュアップを行ない、新規企画の発展を実現した。

ワークショップの参加者・抄録採択者の両者から高い満足度評価が得られ、今後の臨床研究デザインのワークショップのモデルとなることが示唆された。

E. 結論

本ワークショップのような、実際の学会

抄録を用いる方法は、臨床研究のリサーチ・クエスションの構造化や基本事項を学ぶ学習効果を飛躍的に向上させると考えられた。

F. 研究発表

1. 論文発表

特になし

2. 学会発表

特になし

G. 知的財産権の出願・登録状況（予定を含む）

1. 特許取得

2. 実用新案登録

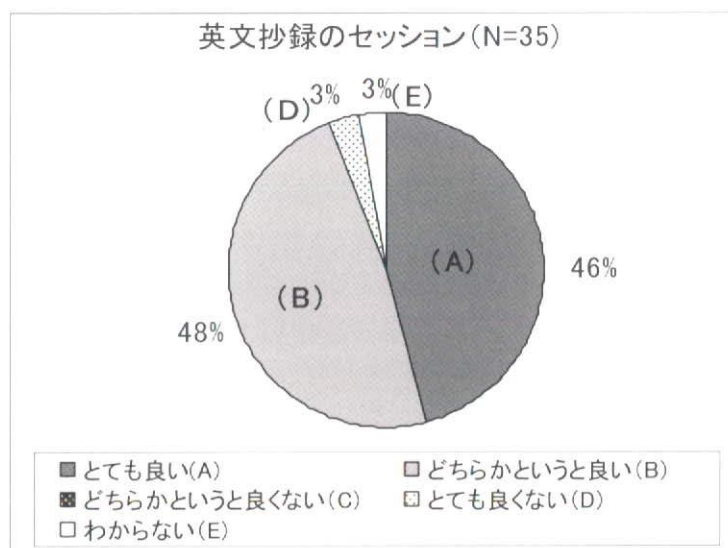
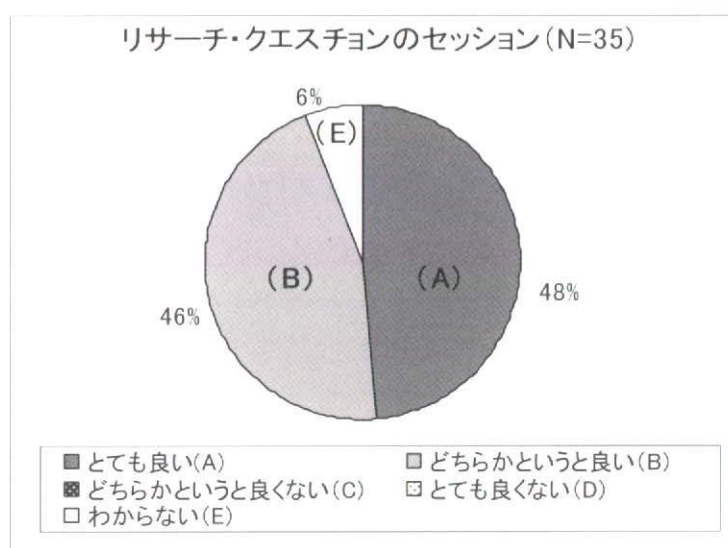
3. その他

特になし

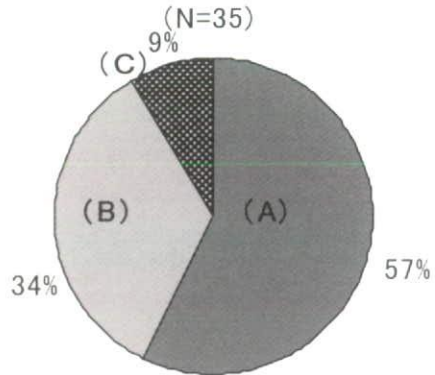


資料：参加者による評価

- ・ 参加者 38 名中でアンケート回答が得られたものは 35 名であった。
- ・ リサーチ・クエスチョン、英文抄録のセッション、グループワーク、全体、全体のバランスに対する満足度に対しては、90%以上の回答者が「とても良い」もしくは「どちらかというが良い」と回答しており、高い満足度が得られた（下図）。
- ・ リサーチ・クエスチョン、英文抄録のセッション、全体の難易度については、セッション全体については 90%以上が適切であると回答した。しかし、英文抄録のセッションについては難しすぎたと回答した人が 23%おり、英語でのレクチャーについては日本語訳をしたハンドアウトを配布する等の対応が必要であることが示唆された。

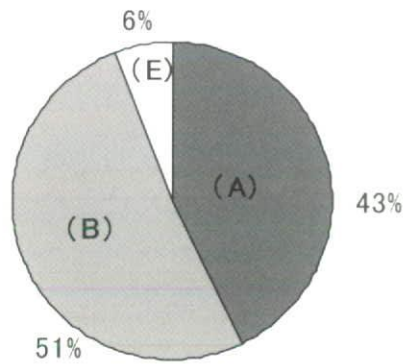


リサーチ・クエスチョンのグループワーク



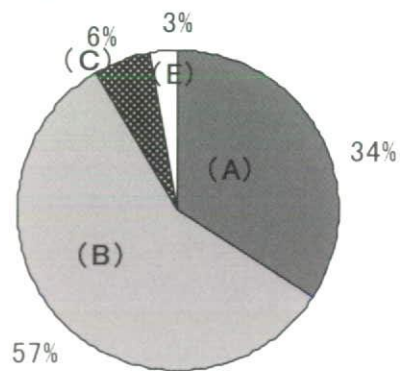
- とても良い(A)
- どちらかというが良い(B)
- どちらかというと良くない(C)
- とても良くない(D)
- わからない(E)

セッション全体 (N=35)



- とても良い(A)
- どちらかというが良い(B)
- どちらかというと良くない(C)
- とても良くない(D)
- わからない(E)

全体のバランス (N=35)



- とても良い(A)
- どちらかというが良い(B)
- どちらかというと良くない(C)
- とても良くない(D)
- わからない(E)

参加者のコメント（アンケート自由記載より）

Awareness（臨床研究への関心）

- ・ 今年から臨床研究についての勉強を始めたところで、その introduction としてよかったと思う。少なくとも今後必要な部分がいくつか分かった。
- ・ 一日大変勉強になりモチベーションがとても高まりました。ありがとうございます。

Awareness（臨床研究の実施）

- ・ 臨床研究は必要とされているが、専門的な教育がない状態で個人の Dr がやっている。統計も不十分な知識で誤ったメッセージが送られる可能性あり、教育は必要と思う。個人的には、社会人入学（夜間や週末に通うスタイル）で受講できればと思います。
- ・ 英語抄録（論文）を書く前に研究デザインがしっかりしていないとダメなことを再度認識できました。ありがとうございました。
- ・ 英語のブラッシュアップでおっしゃったことはわかるのですが、まず英語以前によく内容をつめないといけないということがよくわかりました。海外と日本との違いということもお聞きしたいです。

Education（教育ニーズ）

- ・ 2日にかけての講習会を希望します（土日）
- ・ 定期的に（例えば年に3、4回）あると、とても勉強になると思います。ぜひまた参加したいです。
- ・ 大学院入学（研究開始）後1、2年目の早期だったら良かったと思います。募集時に対象がわかりにくく、当科では高学年だけが参加しました。

ワークショップ全体について

- ・ 参加者の背景が不明ですが、①午前だけでPECO/PICOと概念を行いました、理解度は個人差があったと思われます。（グループワークでも感じました）②ただしAMとPMがつながっていたので、良かったです。（AMに行った実習をPMにさらに深める）
- ・ 開催日、時間的にも適当と思われたが、時間あたりの内容が多く、集中力がきれないためにも休憩をはさんで欲しかった。
- ・ 研究倫理に関する詳しい講義もあればと思います。

リサーチ・クエスチョンのセッションについて

- ・ 午前のセッションをもう少し詳しく時間をかけてやりたかったと思いました。ありがとうございます。
- ・ 午前のセッションでは統計学的用語を普通に使われていて初心者の私には慣れるのが精一杯だった。

グループワークについて

- ・ グループワークの内容は自由度が高すぎて、初対面の人同士で短時間で結論を出すのに苦労しました。もう少し「制限」を加えた方がいいと思います。
- ・ グループワークでやったことの評価をもう少し欲しい。できればその場でアドバイスしてもらいたい。
- ・ グループワークは良かったが、初対面の人ばかりなのでもう少し時間が必要だと感じた。
- ・ 午前のグループレッスンの題材ですが、当グループに abstract を提出していただいた参加者がいたため余計そう感じたのかもしれませんが、結果や考察のない abstractの方がより discussion しやすかったかもしれません。

英文抄録のセッションについて

- ・ 英語で具体的にやっていくのは非常によいです。班ごとの実習も良いですね。
- ・ 英語の講義は少し難しかった。
- ・ 4 大誌によく出てくる word はなぜ出てくるのかという点が重要だと思います。なぜ p-value ではなくて、confidence interval なのかが重要であり、統計的に多いから重要ではないはず。時間的制限があるとは思いますが、わざわざ英語で行うほどの内容でもないと思います。
- ・ とてもおもしろかったですが、自分の英語力のなさで（特に hearing）せつかくのよい講義でも理解が十分でない部分があったので、補足の日本語などあったらよかったです。
- ・ 日本人が誤りやすい英語表現、それらを訂正したものなど、具体例をたくさん挙げてもらいたかったです。



IV. 研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

Shimada T, Noguchi T, Jeffrey L. Jackson, Miyashita J, Hayashino Y, Kamiya T, Yamazaki S, Matsumura T, Fukuhara S. Systematic review and meta-analysis: Urinary antigen tests for Legionellosis. *CHEST* 2009 (in press)

Yamazaki S, Shima M, Ando M, Nitta H. Modified effect of age on the association between ambient ozone and primary care visits at night due to asthma attack. *Journal of Epidemiology* .2009 (in press)

Brazier J, Fukuhara S, Roberts J, Yamamoto Y, Ikeda S, Doherty J, Kurokawa K. Estimating a preference-based index from the Japanese SF-36. *Journal of Clinical Epidemiology*. 2009 (in press)

Yamamoto Y, Hayashino Y, Higashi T, Matsui M, Yamazaki S, Takegami M, Miyachi Y, and Fukuhara S. Keeping vulnerable elderly patients free from pressure ulcer is associated with high caregiver burden in informal caregivers. *Journal of Evaluation in Clinical Practice*. 2008 (in press)

Yamamoto Y, Hayashino Y, Yamazaki S, Akiba T, Akizawa T, Asano Y, Saito A, Kurokawa K, J-DOPPS research group, Miyachi Y, Fukuhara S. Depressive symptoms predict the future risk of severe pruritus in hemodialysis patients: Japan Dialysis Outcomes and Practice Patterns Study. *British Journal of Dermatology*. 2008 (in press)

Utsugi-Ozaki M, Bito S, Matsumura S, Hayashino Y, Fukuhara S, for the MEMO-J Study Group. Physician Job Satisfaction and Quality of Care among Hospital Employed Physicians in Japan. *Journal of General Internal Medicine*. 2008 (in press)

Takegami M, Suzukamo Y, Wakita T, Noguchi H, Chin K, Kadotani H, Inoue Y, Oka Y, Nakamura T, Green J, Johns MW, Fukuhara S. Development of a Japanese version of the Epworth Sleepiness Scale (JESS) based on Item Response Theory. *Sleep medicine*.2008 (in press)

Mishina H, Hayashino Y, Fukuhara S. Test performance of two-question screening for

postpartum depressive symptoms. *Pediatrics International* 2008; 78(3): 307-13.

Miyashita M, Narita Y, Sakamoto A, Kawada N, Akiyama M, Kayama M, Suzukamo Y, Fukuhara S. Care burden and depression in caregivers caring for patients with intractable neurological diseases at home in Japan.

Journal of the Neurological Sciences. 2009; 276: 148-52.

Sugioka T, Hayashino Y, Konno S, Kikuchi S, Fukuhara S. Predictive value of self-reported patient information for the identification of lumbar spinal stenosis. *Family Practice*. 2008; 25(4): 237-44.

Ozaki M, Bito S, Matsumura S. Development of the Japan Hospital Physicians Satisfaction Scale. *International Journal of Health Care Quality Assurance*. 2008; 21(5): 517-28.

Nishiuchi T, Hayashino Y, Fukuhara S, Iwami T, Hayashi Y, et al. Survival rate and factors associated with one-month survival of witnessed out-of-hospital cardiac arrest of cardiac origin with ventricular fibrillation and pulseless ventricular tachycardia: The Utstein Osaka Project. *Resuscitation*. 2008; 78(3): 307-13

Yamazaki S, Fukuhara S, Green J, Takahashi O, Shimbo T, Endo H, Hinohara S, Fukui T. Headache, mental health, and use of medical resources: health diary study in Japan. *Journal of Health Science*. 2008; 54(1): 30-6.

三品 瞳, 三品 浩基, 森田 優治.

「産院小児科外来における個別乳児健診の受診回数および受診時期についての検討」
小児保健研究 2009; 68(3)

三品 浩基, 横山 葉子, 川上 浩司, 福原 俊一.

「臨床医を対象とした臨床研究への関心および教育のあり方についての調査」*医学教育*.
2009; 40(2)

有村保次, 山崎新, 白濱知広, 松倉茂, 千代反田晋, 中里雅光, 福原俊一.

「健康診断における COPD 質問票の有用性の検討」*日本呼吸器学会雑誌*.
2008; 46: 693-9.

前野哲博, 中村明澄, 前野貴美, 小崎真規子, 木村琢磨, 富田絵梨子, 笹原信一郎, 松崎一葉. 「新臨床研修制度における研修医のストレス」 *医学教育*. 2008; 39(3): 175-82.

小崎真規子, 福原俊一. 「病院勤務医の仕事満足度と職場異動希望および臨床からの離脱希望」 *日本医療・病院管理学会誌*. 2008; 45(2): 115-22.

林野泰明, 福原俊一, RESPEQT 研究グループ: 研修医の大学大学院進学希望は低くない, *醫事新報* No.4422:70-74, 2009.

福原俊一, 渡部一宏: 臨床研究フェロシップ構築に関する研究 (Close Up 最先端の医学と患者・社会をつなぐプロを育てる), *GSK Pharmacist Journal*, 22, 12-14, 2008

渡部一宏: 保険薬局における臨床研究のススメ (Step Up 地域で貢献する薬剤師へ), *Quality Pharmacy*, 18, 6-7, 2008

渡部一宏, 福原俊一: 日常業務から Research Question へ ①臨床研究をはじめよう, *Pharma Tribune*, 1, 17-20, 2009

渡部一宏, 福原俊一: 日常業務から Research Question へ ②研究のタネをみつけよう, *Pharma Tribune*, 2, 15-21, 2009

名郷直樹: 臨床研究の ABC EBM の実践と臨床研究を対応させて, *地域医学* (0914-4277)22 巻 4 号:331-334, 2008

名郷直樹: 臨床研究の ABC 臨床現場での情報収集 カルテの情報を臨床研究に生かす, *地域医学*(0914-4277)22 巻 5 号 :441-444, 2008

名郷直樹: 臨床研究の ABC 臨床現場での情報収集 日常診療における観察, *地域医学*(0914-4277)22 巻 6 号 :540-543, 2008

名郷直樹: 臨床研究の ABC 既存のデータを利用した情報収集 厚生労働省統計表データベースを使ってみる, *地域医学*(0914-4277)22 巻 7 号 :664-668, 2008

V. 研究成果の刊行物・別刷

教育 研修 講演

腰椎疾患の臨床疫学研究

杉岡 隆 福原 俊一

京都大学大学院医学研究科医療疫学分野

Key words : 腰椎疾患 (lumbar spine disease), 臨床疫学 (clinical epidemiology), 根拠に基づいた医療 (evidence-based medicine)

臨床疫学とは：違和感から必須知識へ

臨床医学と疫学は一見まったく畑の異なる領域のように感じられるため、“臨床疫学”という言葉には違和感を覚えるものが多く、筆者自身もそうであった。Fletcherらは臨床疫学を「ある特定の患者集団における臨床的問題を科学的に厳密な方法で測定、解析し、その結果から個々の患者についての臨床予測を行う科学」と定義している⁶⁾。ここでいう科学的に厳密な方法とは主に疫学原理を応用した手法を指している。つまりこの定義によれば、臨床的問題に対して現時点で得られた最良の研究成果に基づいて臨床決断を行うという意味での臨床医学と、これら臨床的問題の解決に貢献する研究のほとんどが疫学原理を応用した手法を用いて行われているという意味での疫学が“臨床疫学”という言葉の中に含まれていることがわかる。さらに臨床疫学では、個々の患者の診療を患者が含まれる母集団という枠組みの中で考えたり個々の患者における臨床予測を確率によって表現したりするが、この点でも臨床医学と疫学とが共存している。このように考えると臨床医学と疫学との関係は畑が異なるどころか、実は意外に相補的で相性が良いことがわかる。

最近では常識的な用語となった「根拠に基づいた医療 (evidence-based-medicine : EBM)」も患者診療に臨床疫学を適用するために提唱されたコンセプトである¹⁰⁾。臨床上の疑問を定式化し、それらの疑問に関する最も的確な臨床研究論文を検索し、その情報が臨床決断の根拠になりうるかどうかを病態生理学

的な視点からだけでなく臨床疫学的な視点からも吟味し、そしてそのエビデンスを患者診療に適用する。この一連の作業はこれまでの臨床行為において少なくとも意識的・系統的には行われておらず、これを標準化しようという動きがEBMであるといえよう。したがってEBMを実践するためには臨床疫学に関する知識が必須となる。当初違和感のあった“臨床疫学”はEBMというコンセプトによって今やすべての臨床家にとって必須の知識にまで登りつめたのである。ただし、疫学的アプローチを行いつつ個々の患者の価値観や状況を考慮したうえで最終判断を行うのはあくまで臨床医であって、臨床医の技量が最も重要視される要素であることが従前と変わりが無いのはもちろんのことである。

臨床疫学研究のアウトカム指標

臨床疫学では“アウトカム”という用語が頻繁に用いられる。もともとアウトカムという言葉は日本語で結果・転帰などと訳され、臨床におけるアウトカムとは患者自身の有する何らかの要因や治療行為などの医療的介入がもたらした結果を意味する⁵⁾。そうであればわざわざ“アウトカム”とわずに“結果”と言えば良いことであるが、臨床疫学において最も重要視される臨床的アウトカムは、“死亡するかどうか”、“症状が良くなるかどうか”、さらには“仕事や日常生活に障害をきたすかどうか”、“どれだけ経済的負担がかかるか”などの、患者やその家族にとって切実な問題に関する事柄であり、検査データ

Clinical Research in Lumbar Spinal Diseases

Takashi SUGIOKA et al : Department of Epidemiology and Healthcare Research, Graduate School of Medicine and Public Health, Kyoto University

表1 臨床上の疑問

問題	疑問
異常	患者は病気なのかそうではないのか?
診断	病気を診断するために用いる検査はどのくらい正確なのか?
頻度	病気はどれくらいの頻度で起こるのか?
リスク	病気の可能性を高める因子は何か?
予後	病気の結果として何が起こるのか?
治療	治療によって病気の経過はどのように変わるのか?
予防	健康人に何か介入することで病気を予防できるのか? 病気を早期に発見し、治療することで、病気の経過を改善できるのか?
原因	何が病気を起こすのか? 病気の発症機序は何か?
費用	病気の治療にどのくらい費用がかかるのか?

表2 臨床疫学研究の分類

1. 治療・予防法の有効性・安全性の評価
2. 記述疫学研究(疾患の頻度、分布、自然歴、診療パターン)
3. 要因と害(あるいは益)との関連性を分析する研究
4. 診断法の評価

や病理組織所見の改善ではないことのほうが多い。骨粗鬆症の治療薬によって仮に骨量が検査上あるいは病理学上改善したとしても、骨折頻度が実際に改善しなければそれは患者にとって意味のある治療とは言えない。“結果”と言わずに“アウトカム”という言葉を取って用いるのには、それがこれまで主に用いられてきた、いわゆる“病態生理学的な結果”だけではないということを強調する意図が込められているともいえよう。生身の人間である患者にとって真に意味のあるアウトカムが、医療的介入によって実際に得られているかどうかを直接研究するのが臨床疫学研究であり、また研究結果がより良いアウトカムを得られる方向に臨床や政策を変えていくことが臨床疫学研究の主眼である。

臨床疫学研究の分類とデザイン

Fletcher は臨床疫学で取り扱う主な臨床上の疑問を表1にまとめた⁶⁾。これらの疑問を解決するための臨床疫学研究の種類について、福原は表2のように大きく4つに分類している⁹⁾。また臨床疫学研究のデザインは概ね表3のように分類される。腰椎疾患の領域でもさまざまな臨床疫学研究が行われてい

表3 臨床疫学研究の主なデザイン

I. 観察研究
1. 記述的研究
症例報告や診療実態記述など
2. 分析的研究
a. 横断研究
b. コホート研究
c. 症例対照研究
II. 介入研究
a. ランダム化比較試験
b. 非ランダム化比較試験

るが、以下に臨床疫学研究の4分類に基づいて解説する。なお、本稿では特に観察研究のデザインを中心に、腰椎関連疾患での研究事例を用いながら概説する。

1. 治療・予防法の有効性・安全性の評価

治療や予防法の有効性を評価する場合、通常介入研究が用いられる。ある治療法が本当に有効であるかどうかをみるためには、その治療法以外に結果に影響を与えるすべての要因を除いて判断する必要がある。結果に影響を与える要因としてすでにわかっているものは、それをあらかじめ除外したり解析の時点で調整したりすることは可能であるが、現時点ではまだわかっていないが実は結果に影響を与えている要因については取り除くことができない。唯一それら未知の要因をも平均的に治療群と未治療群それぞれに均等割り付けできる方法が無作為化比較試験である。介入試験の詳述については成書を参考にさせていただきたい。

2. 記述疫学研究

これまで記述疫学研究といえば、疾患の頻度や自然歴、症例報告やその集積などを指していた。しかし近年アウトカム研究の高まりの中で、診療実態や診療のばらつきを記述する研究が増えてきている。比較検討がない単なる現状の記述ということで研究の中では軽視されがちであるが、実は記述的研究こそが次の研究仮説生成のための貴重な情報源であり、原因の探索や診療の改善に大きく貢献する重要な研究である。また逆にいえば記述的研究は単に実態を無秩序に記述するだけではだめで、「いつ、どこで、だれが」といった記述疫学の3要素を踏まえたいうことで現状のどこに問題点があり、その原因は何か、どうすれば改善できそうかといった仮説を想起させるようなものでなければならない。1970年代以降に Wennberg らによって発表された“Practice Pattern Study”は医学・医療界に大きな衝撃を与え、後に1990年代以降のアウトカム研究や診療の質研究の勃興の端緒を切り開くこととなった¹⁴⁾。

診療実態の記述的研究は現在国内ではまだあまりみられないものの、欧米においては例えば腰痛だけを取っても複数の報告がみられる^{14) 11)}。国内における腰痛の記述的研究としては2002年に行われた国民規模の腰痛アウトカム研究がある¹⁶⁾。この研究では20歳から80歳までの日本国民の中から4,500人が無作為抽出され、腰痛の有無や痛みの程度、年齢や職業、他の個人属性、心理社会的要因、包括的健康関連 QOL (SF-36v2 日本語版) などの情報を得ている。その結果、過去1ヵ月間の腰痛の有病割合は全体の30.6%にのぼることがわかり、過去および現在において治療(鍼やマッサージなどを含む)を必要とするほどの腰痛を経験したことのある人においては53.8%と国民の半数以上が腰痛経験者であることがわかった。またこの研究では腰痛の日常生活機能に与えるインパクトを定量化する尺度である Roland-Morris Disability Questionnaire 日本語 JOA 版 (RDQ) の基準値を推定するため年齢・性別ごとに RDQ の平均値や得点分布も算出された²¹⁾。

3. 要因と害(あるいは益)との関連性を分析する研究

1) 横断研究

横断研究ではある一時点において要因やアウトカムの測定を行い、それらの関連性を検討する。この

研究方法は長期にわたって経過を追うことがないので実行しやすいが、逆に要因とアウトカムとを一時点で測るためにどちらが原因でどちらが結果であるかについて言及できないという短所がある。例えば横断研究において腰痛の有無とうつ症状スコアに関連があったとする、しかし果たして「腰痛があるからうつ症状を起こしやすい」のか、あるいは「うつ症状のある人が腰痛を訴えやすい」のか、つまりどちらが原因でどちらが結果かを論じることができない。先ほど例に挙げた腰痛のアウトカム研究では、腰痛あり群と腰痛なし群で SF-36 得点を比較しているが、これは横断研究に該当する。結果は腰痛あり群では「体の痛み」のみならず他の7つの側面においてもおよそ0.5 SD 分だけ腰痛なし群よりも得点が低いことが示され、腰痛というイベントが心身両面の健康へのインパクトと関連していることがわかった。またこの研究では RDQ 値とともに腰痛の痛みの程度を線の長さで表した visual analogue scale (VAS) も同時に測定されており、自覚的な痛みの程度を表す VAS と生活機能障害の程度を表す RDQ との間に乖離があるか、両者に乖離がある場合には心理社会的な要因との関連があるかどうかを横断研究により検討している¹⁷⁾。その結果、痛みが重度でも機能障害の程度は軽いケースや、逆に痛みが軽度なのに障害の程度は重いと答えたケースがいずれも一定数みられ、両者に乖離の存在することが明らかとなった。心理社会的要因との関連については、痛みが軽いが障害の程度は重いと答えた集団は年齢が高くストレスが強うつ傾向があること、さらに時間外労働も多く仕事にも不満を感じているという特性があることがわかり、腰痛診療の際にはこれら心理社会要因を考慮すべきであることが示唆された。

2) コホート研究

“コホート”という言葉は、もともとローマ時代の軍隊における一小隊を意味しており、コホート研究の“コホート”とは共通要素を持つ人間集団を指している。すなわち研究でみようとしているアウトカムがまだ発生していない、共通要素を持った人々の集団を前向きに一定期間にわたってフォローアップするのがコホート研究の方法であり、そのためにもまずは限定された地域や病院などでコホートを設定してベースラインでのデータを得る、その後の追跡に

よりアウトカム発生状況を含むさまざまなデータを随時収集する、というのが基本的な形となる。そしてある要因を有する集団でのアウトカムの発生頻度とその要因を有さない集団でのアウトカム発生頻度とを比較することで、要因とアウトカムとの関連性をみるのが目的となる。コホート研究では前向きに時間を追ってアウトカム発生を調査するため要因とアウトカムとの間には時間的な前後関係がある。したがって横断研究とは異なり両者に関連性がある場合にはその因果関係を論じることが可能で、これはこの研究方法の強みともいえる。しかし長期にわたって多くのデータを収集し続けなければならないため、コホート研究は莫大な経費や人的負担を要する。特に目的とするアウトカム発生の頻度が少ないとさらに長期間の追跡を要するため、できるだけ発生頻度の高いものを対象とするのが現実的である。腰部脊柱管狭窄症に関するコホート研究の例としては、アメリカのMaine州でフォローを受けている腰部脊柱管狭窄症患者を対象に10年以上にわたる追跡調査を行ったChangらの研究がある³¹。彼らは、腰部脊柱管狭窄症に対する手術を受けた集団は手術以外の治療を受けた集団と比べると総じて良好なアウトカムを呈していたが、その効果は時が経つにつれて薄まり、特に痛みや満足度に関しては5年以上経つと差がなくなるという結果を報告している。

3) 症例対照研究

先に述べたように、発生頻度の少ないアウトカムを対象とする場合にコホート研究を用いるのは非現実的である。したがってその場合は、まずすでにアウトカムの発生している人々のデータを集めてその中でもともと要因があった人の割合を調べる。次にアウトカムが発生していない人々のデータを対照群として集めてその中でもともと要因があった人の割合を調べる。そして両者を比較して要因とアウトカム発生との関連性をみる、という形で研究を行う。これを症例対照研究という。この場合注意しなければならないのは、本来ならばコホート研究を行うところをあくまで現実的な方法論として症例対照研究を用いているのであり、したがってできるだけコホート研究に近い形で対象を選定すべきという点である。すなわち症例群や対照群の選定はその背景に共通要素を持ったコホート(母集団)を想定して行われなければならない。例えば、ある地域中核病院の

外来患者での疾患発症例を症例群として選ぶならば、対照群はその疾患にかかったら必ずその病院にかかる(と思われる)人々、具体的にはその病院に別の疾患で普段通院している人や、さらにはその病院の医療圏に在住している健康な人も含めた集団の中から対照を選ぶのが理想である。症例対照研究はコホート研究と比べると時間や経費がかからず実行しやすい点と、要因とアウトカムに時間差があるため両者の関連性から因果関係を論じることが可能である点が強みとなる。しかし実際には対照群の選定を理想的に行うのは困難なことが多く、また要因に関するデータもあいまいになりやすいことから、研究結果に影響を与える情報の偏り(バイアス)を生じやすい。

症例対照研究の実例としてBungeらの脊椎側弯症に関する研究がある³²。これは若年者に対する脊椎側弯症のスクリーニングを評価する研究で、スクリーニングがその後の手術の必要性を減らしているかどうかをみることを目的としているが、手術に至る例が少ないことを考えると症例対照研究はデザインとして妥当な選択といえる。症例群はオランダ全体で脊椎側弯症に対する手術療法の90%以上をカバーしている11病院中の10箇所から選ばれ、対照群は全国37箇所の地域保健サービス機構のデータベースから得られている。この場合ターゲットとして想定されているコホートはオランダ全体の若年者ということになる。著者らはこの研究の結論として、スクリーニングは手術を減らすことに貢献できているとは言えないとしている。

4. 診断法の評価

これまでに紹介した観察研究のデザインはいずれも要因とアウトカムとの因果関係をみることを目的とした研究の事例である。例えば記述的研究は症例や診療実態などの記述からある要因とアウトカムとの関連を探索するのが目的であり、コホート研究や症例対照研究などの分析的研究は主として治療や予防あるいはリスクといった要因が疾患の発生や予後などのアウトカムにどのような影響を与えるかをみるための研究に用いられる。しかし診断に関する研究はある診察法や検査法がどれくらい真の病態に迫れるかをみるのが目的であり、デザイン上もそれらの方法とは異なった形が用いられる。具体的には研究対象である診察法や検査法が、現時点で確立さ

れている診断の至適基準(gold standard)にどのくらい合致しているか、その度合いを表すために、感度、特異度、尤度比といった指標が用いられる(図1)。

1) Gold standard

検査結果の正確度は、疾患の有無に関してその検査で得られた答えが本当の答えとどの程度合致するかで評価されるが、その本当の答えを示す診断方法が gold standard である。例えば X 線上化膿性脊椎炎が疑われた場合、本当に化膿性脊椎炎であるかどうかの答えは脊椎の状態を直接見た時の所見、つまり手術所見が gold standard ということになる。また骨粗鬆症検査における骨塩定量測定の gold standard は骨の病理組織所見ということになるであろう。ただしいずれの疾患も現実にはそこまでの侵襲を加えて真の答えを得るようなことはきわめて例外的である。したがって実際には臨床上現実的にできる範囲の侵襲検査として生検や手術所見、放射線学的手法、あるいは剖検などが、確立された gold standard として用いられている。また、これら侵襲的検査が難しい場合には別の方法が gold standard になることもある。例えば癌のように自然治癒することなく、かつ疑われた時点から数年以内に徴候が顕著になる疾患においては、経過観察の結果が gold standard になる。さらにリウマチ性疾患のように病態が不明で gold standard が存在しないものについては、徴候、症状、特別な検査所見がある数以上持っているという診断基準を満たすことが gold standard になる。この診断基準はエキスパートの話し合いで決定される。

2) 感度・特異度

感度は、実際に疾患を有する者の中で検査が陽性を示す者の割合によって定義される。感度の高い検査は疾患を有する者を見逃すことが少ないので、その検査で陰性が出れば疾患のある確率は非常に低くなる。つまり感度の高い検査を用いることは疾患を除外するのに有効となる。一方特異度は、疾患を有していない者の中で検査が陰性を示す者の割合である。特異度の高い検査は感度の高い検査とは逆で、疾患を有さない者を誤って疾患ありと誤分類することが少ないため、その検査で陽性が出れば疾患がある確率が非常に高くなる。よって特異度の高い検査を用いることは疾患を確定するのに有効となる。なお、実際に疾患があるかどうかは先に述べたように

		疾患		
		あり	なし	
検査	陽性	a	b	a + b
	陰性	c	d	c + d
		a + c	b + d	

$$\text{感度} = a/a + c$$

$$\text{特異度} = d/b + d$$

$$\text{陽性尤度比} = (a/a + c)/(b/b + d)$$

$$\text{陰性尤度比} = (c/a + c)/(d/b + d)$$

図1 感度・特異度・尤度比の求め方

gold standard による診断で決定する。この感度・特異度は検査を受ける前の疾患存在確率によって影響を受けることがないため、これらの指標を得るために行われた研究対象集団の有病割合に関係なく、純粋に検査自体の性能を評価することができる。感度・特異度は陽性・陰性のカットオフ値をどこにするかによって表裏一体の関係で変化する。すなわちカットオフ値を低くすれば見逃しは少なくなるため感度は高くなるが、逆に偽陽性も多く含まれることから特異度は低くなる。逆にカットオフ値を高く設定すると、偽陽性例は減るため特異度は高くなるものの見逃し例が増えるために感度は低くなる。その検査にスクリーニング精度を求めるとかあるいは診断確定能を求めるとかは検査を用いる状況によって変わってくるため、必要な状況に応じてカットオフ値を設定する必要がある。

3) 尤度比

感度や特異度は検査性能をみるためには非常に有用な指標であるが、実臨床の場で検査が陽性、あるいは陰性であったときにその患者に疾患のある可能性がどの程度かを知ることにはつながらない。なぜならその患者が疾患を有するかどうかの確率は検査の結果だけでなく、その検査を受ける前の疾患確率によって大きく影響を受けるからである。尤度比は検査の性能を表すもう一つの指標であるが、検査前確率を用いて検査後の疾患確率を算出することができるため臨床の場でも有用な指標である。陽性尤度比は感度と(1-特異度)との比で検査が陽性であったときの指標、陰性尤度比は(1-感度)と特異度との比で検査が陰性であったときの指標である。陽性

尤度比が高い検査は疾患確定に有効で陰性尤度比が低い検査は疾患の除外に有効となる。ちなみに尤度比を用いた疾患確率の算出にはオッズを使う。オッズとは確率を(1-確率)で割った値で、検査前オッズ×尤度比=検査後オッズという関係がある。また尤度比には感度・特異度でないもう一つの優れた利点がある。感度・特異度を求めるためには検査結果が陽性か陰性かを決定しなければならない。そのためにはカットオフ値を用いて結果を2つのカテゴリーに大きく分ける必要があるが、数値の範囲が広い場合にこれを2つのカテゴリーに大きく分断してしまうのはせっかくの情報を減らすことになってしまう。しかし尤度比の場合は3つ以上のカテゴリーについてもそれぞれ求めることができるため(これを層別尤度比という)、検査の性能をより細かにみる事が可能となる。

腰部脊柱管狭窄症の診断に関する研究

診断法評価研究の実例として、われわれの行った

“腰部脊柱管狭窄症スクリーニング用の自己記入式調査票作成とその検証”について紹介する¹⁵⁾。

1. 研究の目的

本研究は日本脊椎脊髄病学会による『腰部脊柱管狭窄症の診断サポートツール』作成¹²⁾に合わせて行われた。診断サポートツール作成の目的は、プライマリ・ケアを担う一般医が腰部脊柱管狭窄症をほぼ適切に診断できるような簡便なツールを開発することで、腰部脊柱管狭窄症を早期に発見し病状に応じて専門医に相談できる体制がとれるようにすることであった。本研究では腰部脊柱管狭窄症のスクリーニングを、医師の診断を介さずに行えるような自己記入式調査票を開発することを目的とした。この調査票は多忙な地域のプライマリ・ケア外来において整形外科を専門としない一般医が腰部脊柱管狭窄症を見落とさないための補助ツールとしての役割が期待される。また高齢化とともに増加が見込まれる本疾患に関する疫学データ収集や健診などでのセルフスクリーニングにも活用できる。

2. 研究方法

下肢のしびれ・痛みのために参加施設の整形外科外来を調査日に受診したすべての新患者を対象とした。対象患者の選択基準は、①下肢のしびれ・痛みのために参加施設あるいは関連施設を受診した患者、②年齢20歳以上の患者(上限なし)、③本研究に参加することに文書による同意を得た患者、④自力歩行が可能な患者、とした。受診日前1年以内に下肢のしびれ・痛みのために他の専門医療機関を受診した患者(鍼灸は除く)や、認知症・重篤な精神疾患などで判断能力に欠けると推定される患者は研究の

表4 症例の内訳

対象症例	(n=469)	
年齢(平均±SD)	65.2±13.7	
性別(男性)	54.2%	
診断名	n	%
腰部脊柱管狭窄症	222	47.3
腰部椎間板ヘルニア	83	17.7
糖尿病性神経障害	13	2.8
末梢動脈疾患	39	8.3
その他	111	23.7
診断不能例(除外)	1	0.2

表5 選択された7項目とスコア

項目	回帰係数(95%信頼区間)	スコア
年齢		
60歳以上70歳以下	1.21(0.58 to 1.85)	2
70歳より上	1.79(1.19 to 2.39)	3
6ヶ月以上前から症状がある	0.49(0.001 to 0.97)	1
前屈すると症状は改善する	0.97(0.46 to 1.49)	2
後屈すると症状は改善する	-0.99(-1.64 to -0.34)	-2
背筋を伸ばして立っていると症状が起こる	1.22(0.49 to 1.96)	2
歩行時に症状がある場合、休むと改善する	0.57(0.04 to 1.09)	1
尿が漏れることがある	0.63(0.05 to 1.20)	1

対象から除外した。調査実施対象者数は、調査機関(12施設)に集積可能な例数として1施設あたり50名を見積もり、12施設×50名=600名を調査実施対象者数と設定し、実際には469名の対象患者が得られた。この対象患者に対し、性別や年齢、飲酒・喫煙歴、併存疾患などとともに、下肢症状と姿勢(臥位、座位、立位、前・後屈など)や動作(歩行時、自転車運転時)との関連、間欠性跛行の有無、下肢症状の罹病期間や治療歴、下肢の神経症状や膀胱直腸障害の有無など、合計で約100の項目からなる質問紙を作成し、医師の診察を受ける前に配布・回答してもらった。

脊柱管狭窄症の診断方法については診断サポートツール作成と同じであるため、ここでは要点のみ述べる。まずは調査実施機関の医師が患者を診察(病歴・身体所見・MRI)し、脊柱管狭窄症であるかどうかを判断する。次に調査機関に医師が取得したデータ(病歴・身体所見・MRI)を中央パネルに送り、そのデータから中央パネルの医師が脊柱管狭窄症であるかどうかを判断する。両医師の判断が一致した例は診断確定、不一致であった例は後に10名の日本を代表する腰部脊柱管狭窄症に造詣の深いエキスパートによるコンセンサスパネル会議を行い、その合議の結果で最終診断とした。その結果、全469例中1名のみが診断不能で除外された(表4)。腰部脊柱管狭窄症には明確な診断基準が存在しないため、エキスパートが病歴や身体所見、画像所見などから総合的に判断した結果がgold standardになる。本研究ではさらに診断の信頼性を高めるために2人の専門医の診断一致、あるいは10名のエキスパートによる合意を条件としている。

468名のデータを無作為に4:1の割合で分け、374名のデータを調査票開発(derivation set)に、残り

94名のデータを検証(validation set)に用いた。まずは各項目の回答結果と腰部脊柱管狭窄症であるかどうかとの関連を単変量解析で検討した。P値が0.05未満であった22の項目と、P値が0.05以上でも臨床的に有意と考えられる2項目(後屈による下肢症状改善の有無、糖尿病併存の有無)を加えて、ステップワイズロジスティック回帰による変数選択法を用いて項目を選択した。最終的にはP値がすべて0.05未満となる7つの項目が選ばれ、これらの回帰係数の値からスコアリングを行った。回帰係数の最も小さい2項目のスコアを1点とし、これを基準に他のスコアも算出した。

3. 結果

選ばれた7つの項目とそれぞれのスコアを表5に示す。いずれの項目も臨床的に矛盾のないものが選ばれていた。モデル適合度はHosmer-Lemeshowテストで2.47(P=0.9632)と良好な値が得られた。Validation setの94名のデータをスコア順に四分位で分

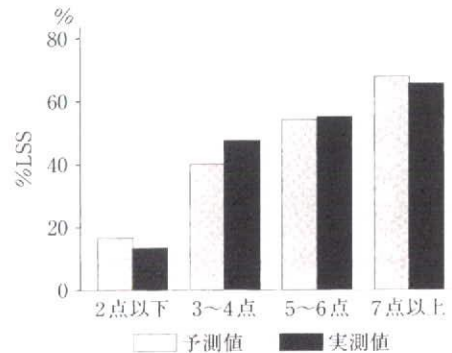


図2 四分位によるスコア群別の腰部脊柱管狭窄症患者割合の予測値と実測値

表6 7項目によるスコアリングモデルの検査性能

指標	Derivation set		Validation set	
ROC 曲線下面積	0.77		0.67	
感度/特異度 (カットオフ: 4点と5点の間)	0.81/0.58		0.75/0.51	
カテゴリー(スコア別)	LSS/total	尤度比	LSS/total	尤度比
2点以下	9/51	0.24	2/15	0.15
3~4点	25/99	0.38	10/21	0.91
5~6点	62/122	1.18	16/29	1.23
7点以上	79/102	3.91	19/29	1.90

けて、各群における脊柱管狭窄症患者割合の予測値と実測値とを比較したのが図2である。予測値と実測値はほぼ一致し、いずれもスコアが高くなるとともに上昇していた。これらの結果からこの7つの項目によるスコアリングモデルは実際の疾患割合をよく反映していることがわかった。

表6はこのスコアリングモデルの検査性能を示している。ROC 曲線下面積は derivation set で0.77, validation set で0.67とまずまずの識別能を示していた。4点と5点の間をカットオフとした場合の感度・特異度は、derivation set で0.81と0.58, validation set で0.75と0.51であった。四分位で分けた層別尤度比は特に低スコア群で低い値となっていた (derivation set で0.24, validation set で0.15)。したがってこのモデルは特に疾患の除外に特に有用であることが示された。プライマリ・ケア外来や健診の場では、腰部脊柱管狭窄症を見逃さないこと、つまり診断確定能よりも疾患の除外能のほうが検査性能として求められる。本研究で開発された調査票はそういった場でのスクリーニングに適した特性を有していることがわかった。今後別の集団に対してこの調査票を用いた検証研究を行い、どれだけのパフォーマンスが得られるか、またこの調査票を用いることで実際にアウトカムは改善されるかをみていく必要がある。

結 語

本稿では臨床疫学研究について腰椎疾患での研究事例を用いて解説した。臨床疫学研究の最終目標は、患者アウトカムが改善するように医師の診療行動に変化を与えること、あるいは医療政策を変えることである。「医学の最新の成果を、速く・安全に・適切に(有効な治療法を必要な人だけに)・安価に(効率よく)、患者・国民の手元に届ける」のが医療のミッションであるとすれば、基礎実験や高度医療機関での狭義の臨床試験だけではそのミッションを達成することは難しい。さまざまな患者のいる一般臨床の場、すなわちリアルワールドでのエビデンスが得られるような研究や、そのエビデンスと実臨床とのギャップ、「エビデンス—診療ギャップ」の実態を調べ、改善に導くような研究が必要となってくる^{8,13)}。また研究の目的が医療のミッションを達成することにあるという意味からも、やはり臨床疫学研究は臨

床家こそが中心になって行うべき研究であると考えられる。もっと言えば臨床疫学研究は一部の専門家だけではなくすべての臨床家が一度は行う、あるいは少なくとも一度は参加するべきものといえよう。今後国内で多くの臨床家が臨床疫学研究に関わり、患者にとって真に必要なアウトカム改善をもたらす多くの成果が世界に発信されることを期待したい。

文献

- 1) Atlas SJ, Wasiak R, van den Ancker M et al : Primary care involvement and outcomes of care in patients with a workers' compensation claim for back pain. *Spine* 29 : 1041-1048, 2004.
- 2) Bunge EM, Juttman RE, van Biezen FC et al : Estimating the effectiveness of screening for scoliosis : a case-control study. *Pediatrics* 121 : 9-14, 2008.
- 3) Chang Y, Singer DE, Wu YA et al : The effect of surgical and nonsurgical treatment on longitudinal outcomes of lumbar spinal stenosis over 10 years. *J Am Geriatr Soc* 53 : 785-792, 2005.
- 4) Côté P, Baldwin ML, Johnson WG : Early patterns of care for occupational back pain. *Spine* 30 : 581-587, 2005.
- 5) Donavedian A 著, 東 高弘訳 : 医療の質の定義と評価方法. 健康医療評価研究機構 (iHope), 京都, 2007.
- 6) Fletcher RH, Fletcher SW : 臨床疫学—EBM 実践のための必須知識—. 第2版, メディカルサイエンスインターナショナル, 東京, 2006.
- 7) 福原俊一編著 : RDQ (Roland-Morris Disability Questionnaire) 日本語 JOA 版マニュアル. 医療文化社, 東京, 2004.
- 8) 福原俊一編 : 臨床研究の新しい潮流—医学研究のパラダイムシフト. 医歯薬出版, 東京, 2008.
- 9) 福原俊一著 : リサーチ・クエスチョンの作り方. 健康医療評価研究機構 (iHope), 京都, 2008.
- 10) Haynes RB, Devereaux PJ, Guyatt GH : Physicians' and patients' choices in evidence based practice. *BMJ* 324 : 1350, 2002.
- 11) IJzelenberg W, Burdorf A : Patterns of care for low back pain in a working population. *Spine* 29 : 1362-1368, 2004.
- 12) Konno S, Hayashino Y, Fukuhara S et al : Development of a clinical diagnosis support tool to identify patients with lumbar spinal stenosis. *Eur Spine J* 16 : 1951-1957, 2007.
- 13) 厚生労働省厚生科学研究 臨床研究基盤整備推進事業 臨床研究フェロロシップ構築に関する研究 2008「臨床研究イントロダクション」<http://www.crfellowship.net/modules/tinyd0/>
- 14) McPherson K, Wennberg JE, Hovind OB et al : Small-area variations in the use of common surgical procedures : an international comparison of New England, England, and Norway. *N Engl J Med* 307 : 1310-1314,