

# アカデミア創生塾

## Summary

4カ月間、土曜 18 時～21 時、全 10 回（月 2～3 回）で、実際の臨床データを使いながら Stata\* という統計ソフトを使って、臨床研究をデザインし、データを解析し、学会・論文発表用の図表の作成方法を学び、その意味を考察できるように修練を積むコース。

\*<http://www.stata.com/>

## My motivation

私が統計ソフト「Stata」を初めて知ったのは 1999 年、Harvard School of Public Health (HSPH) に入学したときでした。それまでは小児科医として多くの時間をがんの子ども達の診療と実験研究に費やしていました。骨髄移植や大量化学療法などにより昔では治らなかった病気を治すことができ、それは医師としての生きがいでした。その反面、目の前で命を落としていった子ども達が多くいたのも事実です。ある患児は移植後拒絶反応を起こし、他の患児は起こさない。ある患児では移植後造血機能の回復が速やかなのに、他の患児では極めて遅く重篤な合併症を起こす。臨床経験を重ねるに従って「一旦発生すると、どんなに一生懸命医療を行っても助けられない病気・合併症がある。病気や合併症の発生リスクを科学的に予測し、これを予防していくことが大切なのではないか？」と強く感じるようになりました。そして「疫学・統計学を 1 から学び直そう」と決心したのです。Stata は HSPH の公式ソフトであり、はじめて触れたときの感動は今でも鮮明な記憶として残っています。同時に「こんなに簡単であれば、もっと多くの人に知ってもらいたい」とも感じました。その後 15 年、やっとその機会に恵まれ、「万感胸に迫る」おもいです。

## Module

1. Ice Breaking Session: Raise your motivation!
2. How to use Stata
3. Designs of Clinical Studies
4. Element of Statistics and Models
5. Basics of Linear Regression Models, using data from [global health](#)
6. Basics of Logistic Regression Models, using data from [cerebral aneurysms](#)
7. How to use “EpiTab” key, using [food poisoning](#) and [medical malpractice](#)
8. Poisson model for [risk minimization plan](#)
9. Practice of [Double blind randomized controlled clinical trial](#)
10. [Repeated Measure](#) of ANOVA
11. Conditional logistic regression model using matched case-control analysis of [SIDS](#)
12. Kaplan-Meier survival curve and Cox proportional hazard model
13. Sample size calculation
14. Multiple imputations, using [birth cohort](#) data
15. Matching with Propensity Score
16. Application of Bayes for diagnosis
17. Mathematical Models for Infectious Diseases, using SARS data
18. Meta-analysis, using VIOXX
19. What was wrong in Jikei Heart Study?

このセミナーは以下の点に工夫を凝らしました。

## **1. 即現場で活かせる**

このセミナーは「即現場で活かせる」ことを意識しています。ほとんどの統計学のセミナーは数式や理論に重点を置いています。このセミナーでは理論を最小化し、各論で実際にデータ解析を何度も繰り返すことに重点を置きました。実践（各論）をやりながら、理論（総論）に戻るといった読み進み方もあると思います。演繹法というよりは帰納法による構成というわけです。

## **2. 実際の臨床データを使って演習できる**

Module 1 ~ 4 は総論ですが、5 ~ 19 は各論です。各論データの多くは、私が関与した臨床研究を教育用に改変したものです。これらのデータを CDROM で参加者にお渡しします。ファイルは①Excel の data file および Stata の ②data file, ③log file, ④do file 各種準備しています。

## **3. 総論**

Stata の使い方、臨床研究および統計解析のエレメントの 3 部で構成されています。総論のカバーする内容は、あくまで各論を理解する上での必要最低限に絞りました。

## **4. 各論**

臨床研究を 16 の類型に分けました。逆に多くの臨床研究はこの 16 類型のどれかに当てはめることができます。本セミナーで Stata の使い方を習得した暁には、臨床研究のアイデアを思いついたとき、「これは第〇章にある研究デザインと統計モデルでいこう」というふうになるはずですよ。

## **5. 各論の構成**

各論の章立ては以下の流れで統一しました。1. プロローグ：研究の背景、2. ミッション：Stata による統計解析の目的、3. (1) デザイン&モデル：臨床研究デザインおよび統計解析のエレメントのどれに相当するかを説明、4. データ：東京図書ホームページ上の Excel file 名、(1) 変数の説明：Excel file の 1 行目にある個々の変数が何を意味しているかを解説し、最初の対象についても具体的に説明、5. アナリシス：Stata による解説箇所各章の中核を成す、(1) 全体の流れ：解析全体の流れを各章ごとに示す、6. エピローグ：プロローグおよびミッションに対する答えを示す。コマンドは四角で囲って見やすくしました。

## **6. 本書内でのリンク**

総論の内容で各論にリンクする箇所、逆に各論の中で総論とリンクする箇所には右上に小さく[ ] で示してあります。[3.2] であれば、3 章 2 節を参照といった具合です。また、( ) で示した数字はその章末の文献番号です。

## **7. コラム**

余力のある人のために設けました。

## **Global health**

この章で強調したかったのは「公開されているデータでさえも **Stata** で解析するといろいろなことが見えてくる」点です。インターネットにアクセスして情報の入手が容易になった現代、情報を知っていることではなく、情報の裏にある真理を分析できることが重要なのだと思います。

## **Cerebral aneurysms**

慈恵医大・脳神経外科と共同で脳動脈瘤の大規模コホート研究を実施しています。1施設で、これだけ多くのデータを管理できることは多くありません。現場医師らの並々ならぬ苦勞が忍ばれます。現在、診断日や最終外来日などの時間的要素を加え、近い将来まとめる予定です。本書では、このデータを教育用に改変して使用しました。

## **Food poisoning**

2011年5月、ドイツで病原性大腸菌 O104 のアウトブレイクが発生しました。あのドイツにおいて 3,816 人の出血性腸炎、845 人の溶血性尿毒症候群、54 人の死亡をだしたことは著者にとって大きな衝撃でした。初期のケース・コントロール研究ではトマト、キュウリ、緑色野菜が原因と発表してしまったため、スペインからのキュウリが原因という噂が流れ、輸入停止にまで発展してしまいました。しかし、その後たった1つのレストランにおけるコホート研究によりスプラウトが唯一の原因と判りました。スプラウトを食べなかった人で発症した人は誰もいなかったからです。しかし「時遅し」で流行は既に終息しつつありました。バイアスを含まないよう研究デザインすることがいかに重要かをよく物語っているエピソードです。この2つの章は、そのような事例を意識して作成しました。

## **Medical malpractice**

19世紀のナイチンゲールの時代から、「死亡率が病院によって異なる」ことが指摘されてきました。しかし、入院中死亡の施設間格差の問題は未だにそのメカニズムの詳細が解明さ

れていません。実態がベールにつつまれ判らない状態であれば、解決策も立てようがありません。ブリストル王立病院のケースは疫学調査により死亡率が有意に高い明らかになり、その後の改善で死亡率が下がっためずらしいケースです。

### **Risk minimization plan**

近年、製薬業界でもリスク最小化策、ベネフィット・リスク・バランスということが重要視されるようになってきました。この事例は、医師が副作用を心配してこの薬を処方しなくなった結果、薬の副作用頻度が減ったというものです。本来であれば「高齢者などハイ・リスクの患者さんらに注意深く使われることによって、副作用が減って然るべし」なのだと思います。データは実データを教育用に改変したものを用いました。

### **Double blind randomized controlled clinical trial**

本研究は我々がパーキンソン病の患者さんを対象に実施した実データを基に、人数を増やすなどして教育用に改変したものを用いました。途中協力医師らが産休や退職等で、登録症例数が伸び悩んだ時期もありました。しかし、何とか論文化できた研究です。研究デザインや統計解析は **Stata** などで解決できます。実際の臨床研究では、リーダーシップ、コミュニケーション、問題解決など **Stata** だけでは解決できない面が多々あります。

### **Repeated Measure**

本データは、リハビリテーション科のドクターとの共同研究から採っています。発想の原点は「ラットの研究で片方の脳に梗塞を実験的に作ると、半体側が麻痺します。しかし、ラットの麻痺は 10 日で回復します。その回復過程において脳を **functional MRI** で観察すると、障害側周囲と反対側にホットエリアを認めました」というものです。1つの研究が終わると、次の研究のアイデアが浮かびます。この連鎖が研究の面白さかもしれません。

### **SIDS**

著者が医師になって間もないころ、SIDS を経験しました。その後、米国で行われた疫学研究により「うつぶせ寝」は SIDS のリスク因子であることが判りました。米国では予防キャンペーンの後、SIDS の頻度は激減しています。どんなに腕のよい医師でも、一度心肺停止状態で運ばれてきた乳児を助けることはできません。将に予防医学の勝利と言えましょう。

### **Survival Analysis**

著者の友人である外科医から症例報告の論文化を相談されたことがあります。既に手術適応がないほどに進行してしまっていた胃がんの患者さんに抗がん剤を投与したところ、がんがきれいに消えてしまい 10 年以上再発なく元気であるというものです。将に患者さんが新しい医療創生のヒントを発してくれていると思います。本章で扱った事例は胃がんではありませんが、我々が行った頭頸部がん臨床研究の実データを教育用に転用しました。

### **Sample size calculation**

著者らの研究グループは、ビタミンDを使った消化管がん再発予防試験を開始しました。終了まであと数年かかる見通しです。ビタミンDががん少しでも患者さんの生存率を改善すれば、こんなにすばらしいことはないと期待に胸を膨らませています。新しい発見をして、それが世界の標準になれば、大勢の命を助けることができます。臨床研究はそんな魅力的な学問だと思います。

### **Birth Cohort**

この出生コホートは臍帯血中水銀濃度と3歳時点での行動との関係を見るのが目的でした。妊娠中マグロなど魚の摂食頻度と臍帯血中の有機水銀濃度はきれいな正の相関を示していました。しかし、臍帯血中の有機水銀濃度とその後の子どもの行動異常との間に相関関係を見出すことはできませんでした。このデータを教育用に多少アレンジしたものです。

### **Propensity Score**

ある国際シンポジウムで「プロペンシティ・スコア」について講演した際、海外の研究者から「ランダム化試験をやらないと駄目なのではないか？」という意見をもらいました。その場では「ランダム化試験をやりたくてもできないときにプロペンシティ・スコアは良い適応なのではないか」と答えましたが、今でもそのやりとりが心に残っています。特にある治療法を選択する際「見えざる交絡: **invisible confounding** (著者の造語)」が強く影響するからです。

### **Diagnosis**

講談社文庫、柳田邦男著「ガン回廊の朝」を十台の頃に読みました。まだ「がん」がほとんど不治の病だった時代、国立ガンセンターの設立当初の話を追ったドキュメンタリーです。早期胃がんを発見するのに二重造影法がよいのか、内視鏡がよいのかしのぎを削るシーンがあったり、ホルマリン漬けのがん組織を徹底的に調べ上げてがんの進行度分類を創った医師たちの話があったりで、医学生であった私にとって本当に胸が熱くなるような内容ばかりでした。本を読む前は、医師の仕事は患者さんを診療することだけとっていました。しかし、この本を読むことによって、新しい医学は「患者さんを治したい」という情熱と弛まぬ努力により創られてきたことをはじめて知りました。それから数十年後、統計学を使って、自ら臨床研究を企画したり、多くの科の臨床研究を支援したりしながら、新しい医学の創生に少しでも寄与していることを誇りに感じます。

### **SARS**

2002年11月16日、広東省で始まった SARS 流行は、916人の死亡を含む8422人の患者をうみ、2003年7月5日に終息しました。世界の研究機関が協力して SARS ウイルスの遺伝子配列を僅か2週間で決定したことはすばらしかったと思います。しかし、このことにより診断法は確立したものの、ワクチンや治療薬が開発されたわけではありませんでした。結局は早期診断早期隔離という古典的方法が功を奏したのです。

2007年、ロンドンの Imperial College で開催されたセミナー : **Epidemiology & Control of**

Infectious Diseases に参加した際、香港で発生した SARS の実データをもらいました。実データ自体はいじらずに、**Stata** を使って解析する教材に仕上げました。

### **Meta-analysis**

最後の章に相応しい社会派の異色研究となっています。医療は産業界からみるとドル箱です。世界中の患者さんが長期に使う新薬を開発すれば、巨額の富を築くことができます。一方、新しい医療は人命に関わることですから、科学的に評価されなくてはなりません。論文を集めて **Stata** で解析すると、1つ1つの論文からは見えてこなかった真理が浮かび上がります。