

シンポジウム②中医学エビデンス再考

漢方処方を使い分けは 弁証なしで統計学的に 再現できるのか ～月経困難症を例に～

吉野 鉄大

慶應義塾大学医学部 漢方医学センター

慶應義塾大学の吉野と申します。よろしくお願いします。

今回は「エビデンス再考」というシンポジウムのタイトルをいただきましたが、そもそも論として、エビデンスは誰のためのものなのかが、最初にあると思っております。専門家がどういうふうに行っているのかを定量化して、またその評価をしていくことは、それ自体が非常に興味深いテーマだと思います。そのうえで、それを誰のためにやっているのかというと、専門家のためというよりは、伝統医学を懐疑的な目で見ている方たちへの説明のためであったり、漢方・中医学に関して全然わからない方々がどういうふうに行うに上手に処方を使い分けていくのかというところに最終的に帰着すると思えます。

日本で9割以上のお医者さんが「漢方薬を使っている・使ったことがある」という状況のなかで、相当数を占めるはずの漢方を専門としない方々が弁証なしでも処方を使い分けることができると、患者さんがより早く、より短期間で自分に合ったお薬にたどり着けるんじゃないかと考えていますので、今回は、専門家向けの内容というよりは、例えば僕の医学部の同級生で全然漢方に関してわかっていない人たちに、上手に漢方薬を使ってもらうにはどうしたらいいのかということが根本にあると考えていただければと思います。

自動問診システムの構築

われわれは2008年から、つまり並木隆雄先生たちがTIASを始められた頃から「自動問診システム」というものをつくって、患者さんの問診データを集める観察研究を行っております。そのうえで、先ほども岩崎鋼先生がお話されたような、漢方診断の予測ができないかどうか、さらに処方選択が予想できないかを重点テーマとして研究を進めてきております。



図1 自動問診システム

「自動問診システム」というふうに申し上げますが(図1)、そういわれると、なにかPepper君のようなものが出てきて、「あなたの今日の調子はどうですか?」と聞かれて、「ちょっと手足が冷えます」と答えるような、対話式の自動問診をイメージされる方もいらっしゃるかもしれませんが、実際には通常の間診票がただ自動でどんどん出てくるというだけです。残念ながら医師の間診そのものを完全に自動化するというふうにはいたっておりません。しかしながら、漢方薬を選んだり、診断を決めたりするうえでは、「患者さんが今どういう状態なのか」が非常に重要になってくると考えますので、「問診票で今どういう状態なのか」という情報をもとに診断であったり処方の選択を予測することができるかもしれないと考えて開発されたシステムです。

そのうえで、患者さんが漢方外来にいらっしゃるとiPadが出てきて、例えば1つの画面に3つの問診項目が出てきて、それぞれに対して「あなたの症状はどうですか?」ということをどんどん入力していきます。それで、3項目の入力が終わると、右下の「次へ>」というボタンを押せるようになりますので、それで次へ、次へと、どんどん進んでいくこととなります。

それで、だいたい80個ぐらいの問診に答えていただいたうえで、さらに主訴を選んでいただいて、例えば「手足が冷える」と患者さんがおっしゃる場合には、それでは「その手足の冷えはどのぐらいか?」をビジュアルアナログスケールで評価することで、データを蓄積しています。

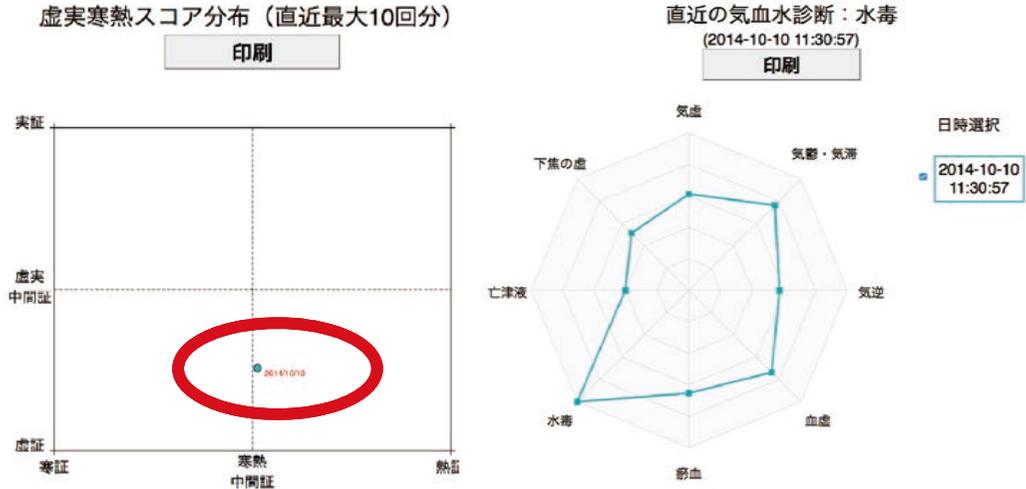


図2 虚実寒熱スコア分析

すでにわれわれがご報告しているとおり、少なくとも虚実と寒熱に関しては、こういった問診情報とバイタル情報・BMI だけでかなり精度の高い予測ができます（図2の左、虚実寒熱スコア分析）。患者さんが入力を終えた段階で、問診情報から機械的に予測した結果を医師も参照できるので「こういった患者さんが次に来るのだな」ということがわかります。そのうえで、専門家はこの予測が正しいかどうかさらにフィードバックをかけて直して式がどんどん変わっていくというようなことをやっております。

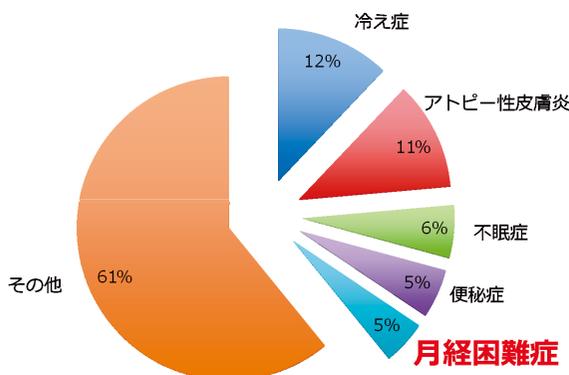
気血水の診断も、寺澤捷年先生のいわゆる「寺澤スコア」の項目のうち、問診でわかるところだけ取り出してスコアをつけて、この気血水のレーダーチャート（同図の右、直近の気血水診断）もつくっており、こちらも、参考程度に表示しています。

こういった研究結果に関しては、千葉大学をはじめ7つの研究施設の方々にご協力いただいて、現状で約8,000人分のデータが集まって、それらを適宜解析しながら進めています。8年半ぐらい続けておりまして、何度かバージョンアップを繰り返しながら、現在もデータを集めている最中です。

■ 月経困難症を例にした自動問診システムによる処方選択の例

ここまで、自動問診システムの紹介を簡単にいたしました。ここから先は処方選択の部分に関して月経困難症を例にご説明します。

例えば当院の漢方医学センター外来では、冷え症が一番多く、その次にアトピーが続きます。こういった疾患に対する処方はどうなのかと気になるのですが、冷え症やアトピーはバックグラウンドがかなり複雑です。老若男女さまざまな方がいらっしゃいます。いろいろなバックグラウンドがあるなかで、さまざまな処方が出ていて、そのままいろいろな処方を予測しようとするとかかなり難しいので



慶應義塾大学漢方医学センター外来2008-2013

図3 慶應義塾大学漢方医学センター外来の主訴の内訳

す。そこで、上から順に見ていって、月経困難症が目につきました（図3）。月経困難症は生殖年齢の女性だけが対象になるため、対象となる処方箋が絞られ、都合がよかったです。

こちらは当院だけのデータですが（図4）、桂枝茯苓丸と当帰芍薬散がそれぞれ3割ぐらいの患者さんに処方されています。残りの4割はそのほかのいろいろな処方が少数ずつでした。そこで、「桂枝茯苓丸と当帰芍薬散だけ」という非常にシンプルなモデルにして「そのどちらがより正しいのか」という予測を行うモデルを作成することにしました。

桂枝茯苓丸と当帰芍薬散は、そのほかにもいろいろと都合のいいところがあります。まずはいずれも月経困難症に対して保険が通っていることです。当院は保険診療施設なので、大学病院の厳しい倫理委員会を通して研究を進めていくうえでは、保険が通らないと観察研究として成り立たないため、保険が通っていることは非常に重要な要素でした。さらに桂枝茯苓丸も当帰芍薬散も、それぞれ日中

処方	人数 (222人中)
桂枝茯苓丸	73
当帰芍薬散	67
加味逍遙散	20
安中散	19
五苓散	18
柴胡桂枝乾姜湯	13
桃核承気湯	10
その他 (50種類以上)	136
のべ処方総数	358

図4 慶應義塾大学漢方医学センター外来全体の処方の割合

韓でRCT（ランダム化比較試験）が行われていてプラセボに対して有意性が示されている処方である点も、研究を進めていくうえで重要でした。そのため「この2つを月経困難症に使い分ける」というようなモデルを進めていくことにいたしました。

まずTSS（当帰芍薬散）群とKBG（桂枝茯苓丸）群の比較をしました（図5）。先ほどの80の問診項目は現在使用している問診項目なのですが、当帰芍薬散と桂枝茯苓丸の使い分けモデルを作成したときのデータは150項目程度の問診を患者さんに入力していただいていたので、患者さんが「その症状があるかないか」を入力したデータを、当帰芍薬散群と桂枝茯苓丸群の2群にまとめてそれぞれ比較し、例えば「桂枝茯苓丸群で有意に頻度が高い項目」といった形で項目を抽出してくると、ご覧の9項目が抽出されました。

問診項目では、当帰芍薬散群は、症状としての「めまい」「痩せ型」（BMIで18.5未満の方）、さらに医師の診察結果の「腹力虚弱」が陽性である頻度が高くみられました。それに対して、桂枝茯苓丸群は、症状としての「腰の冷え」「汗をかきやすい」「暑がり」「足がしびれる」に加えて医師の診察結果の「腹力が充実している」「臍周囲の圧痛」（いわゆる瘀血の抵抗）の頻度が高くみられました。このあたりは専門家にすれば当然の結果が出てきたと考えています。

これらの項目を使って、最終的に、先ほどの岩崎先生のような形で二項ロジスティック回帰分析を行ってモデルをつくることになります。この9項目を候補にしたモデルをつくるとしても、たった1つの項目だけでモデルをつくることもできますし、2項目でも9項目でもいいのです。そうなると2の9乗のマイナス1、つまり511の組合せがありえます。そのなかで、どのモデルが一番いいのかを考えていきます。

まず「ほどよい問診項目数でよい予測を行うにはどうしたらいいのか」を判断するために赤池情報量規準という統計学的な基準を使い、「511の組合せの候補から、7項目を含んだ最適の組合せ」を選んできました。そのうえで、ROC曲線下面積をみて、「7項目を使った最適モデル」が、例えば「9個全部使った場

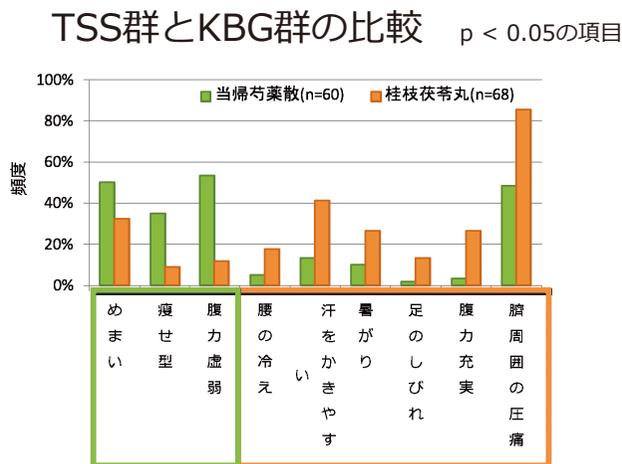


図5 当帰芍薬散群と桂枝茯苓丸群の比較

合のモデルと比べていいかどうか」を比較しました。さらに、ホスマー・レメショウ検定でp値を用いて「7項目を使った最適モデルがうまくいっているかどうか」をみました。さらに、クロスバリデーションというやり方を使って、元のデータセットを2つに分割して、一方で予測モデルをつくり、それをもう一方にあてはめて、予測性能を調べるという検証を行いました。

さらに、先ほどの岩崎先生の報告発表にもあったように、「まったく新しいデータに対するあてはまりはどうか」を検証します。モデルを作成するのに使用したのは2008年の5月から13年の3月までのデータです。その翌月（2013年の4月）から2015年12月までの「新規データ」で、月経困難症で当帰芍薬散もしくは桂枝茯苓丸が処方された患者さんのデータにこの処方予測モデルをあてはめて、「その医師の処方と予測モデルの予測が一致しているかどうか」を検証したところ、8割5分の一致率が得られました。2択としては十分に受け入れられる一致率が得られたのではないかと考えましたので、この画面の右下に出てきているような（図6）、「7項目それぞれがあるかどうか」チョンチョンとチェックすれば「処方のどちらがいいか」を表示するアプリケーションをつくることまでいっております。

しかし「とにかく月経困難症に対して当帰芍薬散か桂枝茯苓丸かを、いつ誰が入力しても同じ結果が出てくるモデルができた」と吉野は言っているけれども、いろいろ限界があるんじゃないかと思われるはずですが、先ほど「7施設のデータがあります」と言っておきながら、今回は残念ながら慶應のデータしか使うことができませんでした。ほかの施設だと月経困難症は非常に頻度の低い病名だったので、ほかの施設のデータを使うと数サンプルしか月経困難症のデータが出てきませんでした。ほかの施設でどうなのかは今後の検討になります。

ほかには、先ほどご質問に出たような、患者側のばらつきや、医師側のばらつきも、あると考えています。岩崎先生は、1人の患者さんを2人で診察することでCohenのカップパ係数を計算して、医師の判断の一致率をみておられますが、われわれはそういったことはやっていません。純粋に日常診療のなかで1人の医

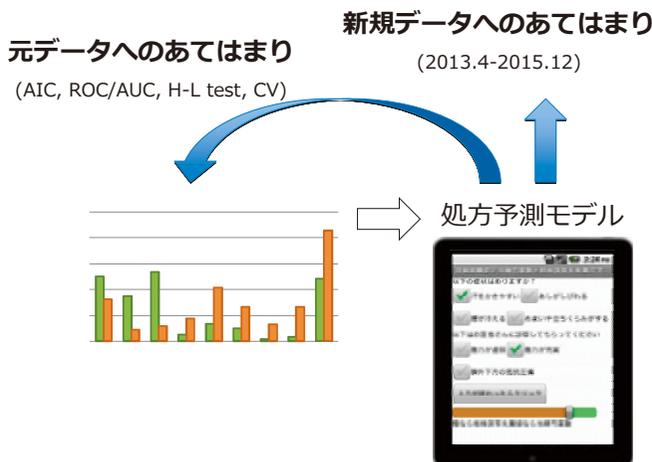


図6 処方予測モデルの検証

師が1人の患者さんを診て、そのデータをそのまま入力していますので、医者同士で考えていることが全然異なったりすると、まったく別の結果が出てくる可能性があります。ただ、最終的に予測自体が8割以上の一致率でうまくいっているので、少なくとも患者側のばらつきであるとか医師側のばらつきに関しては、あまり大きな要素にはなっていないと考えています。

臨床研究デザイン

今回の研究の最大の問題は、あくまで「漢方の専門医が処方をこういうふうを選んでいく」ということが予測されるだけという点です。例えばこのモデルで「当帰芍薬散がいいです」と予測されたとしても、「この患者さんが当帰芍薬散で治るかどうか」はいっさい検証されていません。実際には、われわれはこの200人近くの月経困難症の患者さんの経過をVASで追ったりカルテで追ったりして、おおむね8割以上の患者さんがなにかしらの形で良くなっていることを確認していますので、大部分の方は良くなっていくのだろうとは思いますが、治らない方もいらっしゃるし、副作用を経験する方もいらっしゃいます。そういった方々をどう取り扱うのかといったところに関しては検証する必要があります。最終的に、われわれのこのモデルを用いた介入の有用性について、つまり「実際に患者さんがうまく治っていくのか」「副作用は少ないのか」を前向きに検証していく必要があると考えています。

ということで、今このモデルを使った前向きの臨床研究に関しても倫理申請を行っておりますので、「こんなことを考えています」という点を最後に少しだけ共有して、終了といたします(図7)。

まず、月経困難症の患者さんに対する臨床研究ですので、もちろん月経困難症がある、生殖年齢の女性患者さんを集め、ランダムに2群に割りつけます。予測モデルを使って当帰芍薬散もしくは桂枝茯苓丸を飲んでいただく漢方治療群をコントロール群と比較して、どちらがいいのかを検証します。

試験デザインにおいて、漢方治療群よりも、コントロール群に何を設定するかの方が議論になっています。当帰芍薬散も桂枝茯苓丸もすでにプラセボに対し



図7 臨床研究デザイン

て有意差をもって症状を改善することが示されたお薬なので、今さらコントロール群にプラセボをもってくるのは倫理的に難しいことになりまして、おそらくプラセボを対照にすれば有意差は示せると思うのですが、プラセボを対照とする必要性は少ないと考えています。

この試験で知りたいことが「漢方薬を使い分ける意義」だとすれば、「漢方薬を使い分けられない群と比べた方がいいのではないか」ということになります。例えば「当帰芍薬散を一律に投与する群と、漢方薬を使い分ける群とを比較したらいいんじゃないか」というアイデアもあると思いますが、それでは、両群で半分程度の患者が同じ治療を受けることになってしまうので、それはコントロール群として不十分だと思います。

ほかには「標準治療を受けた群はどうか」ということになります。月経困難症の患者さんで標準治療というと、低用量ピルを飲むかNSAIDsを使うかということになりますが、そもそも標準治療でうまくいくのであれば、それでよいわけですし、漢方外来にいらっしゃる患者さんは痛み止めを飲んででも効果が不十分だったり、副作用などにより「ピルを飲みたくない」という方がほとんどなので、標準治療をコントロール群にするのも難しいと考えます。

そこで、現状われわれが考えているコントロール群は、モデルによる予測の「逆張り」です。帰無仮説として漢方薬を使い分ける意義がないのであれば、われわれがつくった予測モデルで当帰芍薬散と予測された患者さんに「桂枝茯苓丸を出す」という「逆張り」をすれば、それがコントロール群になるはずで、そういった帰無仮説が棄却できれば、「漢方薬を使い分ける意義」が示せるのではないかと考えています。

以上、「漢方処方を使い分けは弁証なしで統計学的に再現できるのか」というテーマを設定してお話させていただきました。われわれのところは、大塚敬節の流れを引き継ぐ渡辺賢治先生のもとでやっておりますので、いわゆる古方派であって、そもそも弁証をしていない方々たちのデータですので、弁証をしていない人たちのデータは弁証なしで統計学的に再現できるというのは当然といえば当然なのですが、こういったシステムが中医学にも応用されていけば、両者が共通言語で語っていくことができるのではないかと考えました。ありがとうございました。