

健康文化

医学文献の評価

二橋 尚志

はじめに

臨床現場では、ますますエビデンス⁽¹⁾に基づく正しい判断・決断が求められるようになってきています。インターネットが発達し、日常臨床において活用できる情報が爆発的に増加し、また、使用できる検査項目、機器なども多様化しつづけ、臨床家は、忙しい日々の診療のなかで、そうした多種多様な情報を駆使し、または、終わることのないアップデートをつづけ、診断、治療を実施していくことが求められる状況があるかと思えます。

診断とは

一般的には、“普通ではない”もしくは“日常生活に支障がある”などの理由で相談を受けた医師が、患者の病歴を詳しく聴取し、臨床徴候を調べ、その後、医師は“〇〇が原因ではないか”“〇〇病かもしれない”というように診断仮説を立てます。この段階で、“感冒”であるとか“帯状疱疹”であるといった自身の仮説が、かなり確からしいとか、現在の仮説に基づき、治療を始めても、患者の不利益はおそらく生じさせず、利益をもたらす可能性が高いと考えられれば、そのまま、治療をするかもしれません、しかし、多くの場合、血液検査、検尿、心電図、画像検査などを自身の仮説に基づき必要なものを選択し、治療前に仮説が正しいか、新しい仮説を立てなおす必要があるか検査を実施します。

診断ガイドライン

次いで、検査を実施しようとした後で、どの検査を選択すべきかといった問題があります。“自分のたてた仮説を効率的、かつ、正しく証明するのに最適な方法を自身が選択できているだろうか？”“もしかして患者に不利益を与えてないだろうか？”といった不安や心配を時に覚えるかもしれません。

そんな時には、“同じような状況で、他の医師達はどのようにしているだろうか？”といったことを、調べると思えます。そうしたときに、ガイドラインなどと言うことが脳裏に浮かんできます。診療ガイドラインは、医療者と患者とが特定

の臨床場面で適切な決断を下せるように支援する目的で、体系的な方法に則って作成された文書と定義され、これは Evidence based medicine に従って作成することとなっています⁽²⁾。強い Evidence、弱い Evidence などを区別し、ガイドラインを作成するわけです。一般的に用いられているものとしては Oxford Centre for Evidence-based Medicine Levels of Evidence があげられます⁽³⁾。これでは、1) Therapy / Prevention, Aetiology / Harm。2) Prognosis。3) Diagnosis。4) Differential diagnosis / symptom prevalence study。5) Economic and decision analyses。といった 5 つの項目に対して階層的な評価を行っており、各項目により、微妙な違いがありますが、強い Evidence はシステマティックレビューや randomized controlled trial ということになっており、一方、弱い Evidence は Expert opinion と記載してあります。

医学文献の評価

いろいろな文献を調べて、その結果を実臨床に応用する場合に、もちろん、強い Evidence と評価されれば、一安心ではありますが、その文献が本当に正しい評価を下しているかどうか、疑問はなかなかゼロにはなりません。

ここで具体的な事案を考えてみます。自分と同じ検査を使用し、脳腫瘍の再発か治療後変化か判定しようといった臨床状況で、研究Aは、感度 90%、特異度 90%と報告しており、一方、研究Bは、感度 70%、特異度 50%と報告しているなどと、同じような検査を実施しているのに、結果が異なっているといった状況はよく遭遇します。それでは、このどちらの結果を採用して、自分の直面した状況にあてはめるか決めなければいけません。研究Aの方を選択したい気はあるのですが、何か基準はないでしょうか？もし研究Aの結果に信頼がおけるものであれば、“この検査をおこなうと、まず、正しく診断できるので是非やりましょう” といえそうです。しかし、研究Bの方が、確からしいということになれば、“診断できない可能性もありますが、他に方法がないから、やってみましょう” など、だいぶ、話すトーンが異なってくるそうです。では、どちらが確からしいか、もしくは自分の臨床状況に近いのかをきめる一助として、Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS) tool といったものが提案されています⁽⁴⁾。読んで字のごとく、診断に関する研究の質を計るための道具で、systematic review を作成する時に、広く汎用される道具です。

これは、大きく 4 つのドメインからなっており、それぞれ low risk、high risk、unclear と三段階の評価を行い、要するに採点をします。

表1. QUADAS2 結果

	Risk of Bias				Applicability Concerns		
	Patient	Index	Reference	Flow and	Patient	Index	Reference
	Selection	Test	Standard	Timing	Selection	Test	Standard
研究 A	⊖	⊕	⊖	?	⊖	⊕	⊕
研究 B	⊕	⊕	⊕	⊕	⊕	⊕	⊕

⊕ : low risk, ⊖ : high risk, ? : unclear

4つのドメインは、1) patient selection (対象となる患者群がきちんと選択されているかなど)、2) index test (検査がきちんと実施されたかどうか。結果の解釈は妥当かどうかなど)、3) reference standard (臨床状況、上記の場合は、再発か治療後変化かがきちんと評価されているかなど) 4) flow and timing (検査は、同じ状況で実施されているか。検査をやっている群とやってない群を含んでいないかなど) からなり、それぞれのドメインは、まず risk of bias として評価され、最初の3つに関しては、concerns regarding applicability として評価されます。

例えば、patient selection の risk of bias の評価として3つの質問が用意してあります。1) Was a consecutive or random sample of patients enrolled? (患者の登録は、連続もしくはランダムですか?) 2) Was a case-control design avoided? (ケースコントロールデザインは避けられていますか?) 3) Did the study avoid inappropriate exclusions? (不適切な除外を避けていますか?)。こうした評価により、点数の低い文献の信頼性には疑問があるということになります。表1のように評価すると、研究Aの報告は Patient selection や Reference standard に問題がありそうです。このような報告に多いですが、再発か治療後変化か、病理診断もしくは経過観察で診断したと書いてあることがほとんどです。病理診断も生検か全摘かで、精度に大きな違いがありそうですし、かりに生検であるとする、目標をどう決めたかも重要です。経過観察であれば、観察期間がどれくらいかは、記載した方が良さそうですし、治療後いつ、評価したかにより結果が違ってくるかもしれません。

研究Bはすべての項目を正しく、かつ公平に評価しているようです。そうすると、研究Aの結果は診断精度が高く魅力的ですが、QUADAS2 で評価すると、問題があるかもしれないので、研究Bの結果を自身の臨床状況に応用した方が

現実的かもしれません。ただ、臨床で遭遇する症例は、同一の疾患であっても、経過など臨床状況は多様であり、なんらかのバイアスは必ずありそうです。そういう状況で、判断すると考えておいた方がいいでしょう。

最後に画像診断に関して

画像診断といっても、様々です。まず、病気があるのかどうかといった“存在診断”、癌か炎症性結節かといった“鑑別診断”、癌であれば、“病期診断”や“悪性度診断”、実際に治療がはじまれば、“治療効果判定”、治療が一通り完了し病気が消失した後では、“再発診断”、また、さまざまなタイミングで実施する“予後予測力”などなど、いろいろあります。この状況に関して、上記の種々多様な診断を下す必要があります。それぞれに関して、日々、医療が進歩し、数多くのエビデンスとして医学文献が発表・蓄積され、そういう状況で、リアルに臨床に対応する医学文献を選択し、すなわち、強いエビデンスを取捨選択していくのは、忙しい日常臨床の中では、大変難しい作業ではないかと思えます。

参考文献

1. 臨床のための EBM 入門 決定版 JAMA ユーザーズガイド. 監訳 古川 壽亮, 山崎 力. 医学書院.
2. Institute of Medicine. Clinical Practice Guidelines: Directions for a New Program. Field MJ and Lohr KN, eds. Washington, D.C.: National Academy Press: 1990.
3. Oxford Centre for Evidence-based Medicine Level of Evidence. (<http://www.cebm.net/index.asp>).
4. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. Ann Intern Med. 2011. 18;155(8):529-36.

(名古屋大学医学部附属病院・助教)