

- 5) ケアの供給者は目隠しされているか
- 6) 患者は目隠しされているか
- 7) 一次エンドポイントの点評価値とばらつきが指標が示されているか
- 8) 治療企図分析が行われているか

IV. Minds における医学文献評価選定基準

現在(財)日本医療機能評価機構¹⁴⁾では Web 上で検索可能な診療ガイドライン情報の提供とともに医学文献の内容を構造化抄録として Web 上に検索できる形で提供する医療情報サービス事業 Medical Information Network Distribution Service 通称 Minds と呼ばれる事業が進行中である。著者は医学文献評価選定委員会委員長を勤めさせていただいているがここで用いる予定の医学文献選定法について紹介する。従来の医学文献の評価法をさらに掘り下げて客観的で信頼できる評価法を追求しようとしている。なおすべての過程が確定しているわけではないので将来変更される可能性があることをお断りしておく。

文献検索はクリニカルクエスチョンからスタートする。検索式は 5 つの項目から構成される。1) 疾患/病態 2) 予知因子 (介入/危険因子) 3) 対照 4) アウトカム 5) 研究デザインであるが検索結果に応じていずれかの項目を削除したり OR で語句を追加したりしながら柔軟に対応する。(表 3)。

検索結果からタイトルとアブストラクトに基づき次の選定基準で選定を行う。

- 1) 言語：英語または日本語。
- 2) 人を対象とした研究。
- 3) 1) 疾患・病態 2) 予知因子 (介入曝露) 3) 対照 4) アウトカム 5) 研究デザインの各項目がクリニカルクエスチョンと一致している。
- 4) 必要に応じて委員会が指定した語句を含むあるいは委員会が指定した条件を満たす
- 5) さらに ACP Journal Club¹⁵⁾ の採択基準を参考に研究目的によって分類しそれぞれについて以下に述べる基準を満たすものを選択する。
 - 1) 予防治療に関する研究 (Prevention or treatment)
 - A) ランダム割付が行なわれている
 - * なおランダム化比較試験以外でも委員会で適格と判断したものは採用する。
 - 2) 診断に関する研究 (Diagnosis)
 - A) 対象疾患だけでなく対照疾患を含む
 - B) sensitivity, specificity, likelihood ratio, positive rate, positivity, odds ratio などの語句を含む

- 3) 予後に関する研究 (Prognosis)
 - A) コホート研究 (なお曝露による群分けをしていない研究も含める)
- 4) 病因に関する研究 (Etiology)
 - A) 曝露と推定されるアウトカムの関係を探求している
 - B) アウトカムに対するリスクがある明確に定義された群で前向きにデータ収集が行なわれている：次の順で望ましい。ランダム化比較試験, 準ランダム化比較試験, 非ランダム化比較試験, 症例ごとにマッチングが行われたか比較する群を生成するのに統計学的調整が行なわれたコホート研究, ネスティッド症例対照研究 (なお症例対照研究も研究デザインの分類は行う。まれな疾患や病態がアウトカムの場合には症例対照研究も採用する)
- 5) 医療の質の向上あるいは持続的教育に関する研究 (Quality improvement or continuing education)
 - A) ランダム割付が行なわれている
- 6) 医療保健プログラムあるいは介入の経済に関する研究 (Economics of health care programs or interventions)
 - A) 日本を対象にしている
 - B) 研究対象の経済的な問題は実際の患者における選択肢の比較に基づいている
- 7) システマティックレビューまたはメタアナリシス (Systematic review article or meta-analysis)
 - A) 特に条件を設定せずすべて採用する。

[注]

・臨床予測ガイドに関する研究 (Clinical prediction guides) については Minds abstract の作成は行わない。文献の存在を紹介するのみにとどめる。

・鑑別診断に関する研究 (Differential diagnosis) は診断に関する研究と同様に取り扱う。

さらに Meaningful すなわち意味があるかどうかすなわち Understandable かどうかで評価したうえで全文を取り寄せる。全文から EBM データテーブル構造化抄録 (Minds abstract) を作成し EBM レビュアーと疾患専門レビュアーのチェックを受け最終的に医学文献評価選

定委員会でチェック、確定した上で Web に掲載される。委員会では 1) Validity (妥当性) 2) Reliability (信頼性) 3) Relevance (臨床的意義) の 3 つの項目について半定量的に評価を行う計画である。Meaningfulness (意味) も含めこれら 4 つの概念は互いに関連する部分もあるが独立した概念である。簡単に解説すると次のようになる。

Meaningfulness は意味を持つかどうかという概念である。言い換えると評価者が理解できるかどうかを問うことを問う。理解できるかどうかは知識経験能力によって異なるので一定の水準を越えた評価者が評価する必要がある。

Validity はそれが測定しようと意図していることを測定しているかどうかを問う概念である。例えばプラセボ対照ランダム化比較試験で癌患者の生存率が改善することを証明しようとする場合に腫瘍縮小効果があることだけを証明してもそれは意図していることを測定していることにはならないので Validity は低い。

Reliability は再現性を問うものである。同じ研究を繰り返した場合同じ結果が得られた場合 Reliability は高い。

Relevance は臨床との関わりの深さであり臨床的意義のことである。患者にベネフィットがあることを直接説明しているかどうかの問題となる。一方でそのベネフィットの大きさも相対的に評価すべきであり特に介入については従来の介入と比較して相対的にその臨床的意義を評価すべきであろう。ベネフィットの大きさはさまざまな効果指標を用いて表すことが出来るが絶対値だけを問題とするのではなく相対的に評価されるべきである。

V. 最後に

医学論文の評価にはさまざまな困難が付きまとう。論文にすべてが記述されているわけではなく記述されていなくても実行されていたり逆に記述されていても実際には不十分にしか実行されていない場合もありうる。研究デザイン以上に細部の評価をすることは多大な労力と時間を要するので研究デザインに基づいた評価を行ないそれにしたがって勧告の強さを決める方法は実用性を考えると受け入れざるを得ない。しかし同じランダム化比較試験であってもさまざまな質のものが存在するのも事実であり今後ここで述べたようなより詳細で精度の高い評

価法が普及する可能性がある。

参考文献

- 1) National Institute for Clinical Excellence [internet]. <http://www.nice.org.uk/> [accessed 2004-2-19]
- 2) Cook DJ, Greengold NL, Ellrodt AG (et al). The relation between systematic reviews and practice guidelines. *Ann Intern Med* 1997;127(3):210-6.
- 3) 福井次矢. EBM に基づいた診療ガイドラインの作り方. *EBM ジャーナル* 2000;1(4):438-43.
- 4) Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? : The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281(20):1900-5.
- 5) The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12(1):18-23.
- 6) The AGREE Collaboration [internet] <http://www.agreecollaboration.org/>
- 7) NICE ガイドライン作成マニュアル [internet] <http://www.nice.org.uk/pdf/guidelinedevelopmentnccsfinal.pdf> [accessed 2004-2-19]
- 8) Zaza S, Wright-De Agüero LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH (et al). Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18 (1suppl): 44-74.
- 9) 森實敏夫. 臨床医のための EBM アップグレード. 東京: 医学書院;2002.
- 10) Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; 52(6):377-84.
- 11) Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287(22):2973-82.
- 12) Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282(11):1054-60.
- 13) Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM (et al). The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; 51(12):1235-41.
- 14) (財)日本医療機能評価機構 [internet] <http://jcqhc.or.jp/> [accessed 2004-2-19]
- 15) ACP Journal Club [internet] http://www.acpj.org/shared/purpose_and_procedure.htm [accessed 2004-2-19]

肝疾患の予後

森實 敏夫



- ▶ 比例ハザード解析に基づく多変量のモデルにより、リスクスコアを算出し個別症例の予後を予測することができる。
- ▶ 原発性胆汁性肝硬変，原発性硬化性胆管炎の Mayo モデルを用いると，臨床症状，臨床検査の結果から予後を予測可能である。
- ▶ 消化管出血のない肝硬変症の予後の予測は，Child-Pugh スコアおよび BUN，平均赤血球容積，%前腕筋周径から可能である。

予後とは

予後を知るために測定される項目，すなわちアウトカムは，急性の疾患であれば“治癒”，慢性の疾患であれば，それぞれ症状の改善および全体としての“QOL” (quality of life)，そして死亡を引き起こす疾患であれば“生存”が最も重要である。

複数の因子によるモデル

時間の経過のなかで起きるアウトカムへの，多くの因子の寄与を解析する方法が，比例ハザード分析である。複数の因子の生存への寄与の程度は，係数で表し，その係数の値を各因子の値にかけ算することによって，重み付けして，それらの値を合計した値を求め，生存期間を予測することができる¹⁾。この合計した値を，“リスクスコア”あるいは“予後指数(スコア)”と呼ぶ。

リスクスコア = 係数₁ × 因子₁ + 係数₂ × 因子₂ + …… 係数_n × 因子_n

リスクスコアが大きな値をとると，予後は悪くなる。すなわち，生存期間が短くなる。逆に，リスクスコアが小さな値をとると，予後は良くなる。すなわち，生存期間が長くなる。したがっ

て，1年後，2年後，などの生存確率を計算することも可能である。

リスクスコア R の場合の t 時間後の生存率 = $\{S_0(t)\}^{\exp(R-R_0)}$

$S_0(t)$ は全症例のリスクスコアの平均値 = R_0 の場合の，t 時間後の生存率で，ベースライン生存率と呼ばれる。exp は exponential のことで，自然対数のべき乗のことである。

上記の各係数の値，ベースライン生存率の値は，論文から得られる。個々の症例の各因子の値を代入して，1年後，2年後，などの生存率を算出すると，その患者の生存曲線を描くことができる。

原発性胆汁性肝硬変

年齢，症状と臨床検査だけからなる Mayo のモデル²⁾(☆☆)が広く使われている。

リスクスコア = 0.871 × 血清総ビリルビン値 (mg/dl) の自然対数 - 2.53 × 血清アルブミン値 (g/dl) の自然対数 + 0.039 × 年齢 (年) + 2.38 × プロトロンビン時間延長 (秒) + 0.859 × 浮腫スコア

浮腫スコア：浮腫なし 0，利尿剤投与なしで浮

もりざね としお：神奈川県横須賀市稲岡町 82 神奈川歯科大学内科 ☎ 238-8580

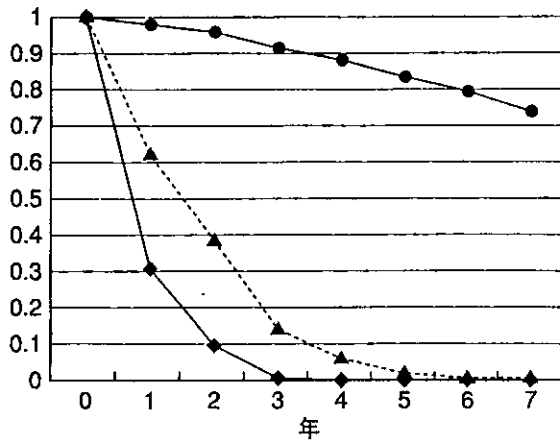


図1 比例ハザードモデルから求めたPBC患者の生存曲線
◆は条件1の例、●が条件2の例。縦軸は生存確率。
▲はリスクスコア7.8の場合。

腫あり0.5,利尿剤投与で浮腫なし0.5,利尿剤投与で浮腫あり1.0, $R_0=5.07$ 。ベースライン生存率:1年0.97,2年0.941,3年0.883,4年0.833,5年0.774,6年0.721,7年0.651。

例えば,かなり進行した例を想定して,血清総ビリルビン値3.0,アルブミン値3.0,年齢65歳,プロトロンビン時間延長3.0秒,浮腫スコア1.0の患者の場合(条件1),
リスクスコア $R=0.871 \times \text{LN}(3.0) - 2.53 \times \text{LN}(3.0) + 0.039 \times 65 + 2.38 \times 3.0 + 0.859 \times 1.0 = 8.71$
となる(LNは自然対数)。 $R - R_0 = 8.71 - 5.07 = 3.64$ となるので,例えば,1年後の生存確率は,
 $0.97^{\text{exp}(3.64)} = 0.31$ となる。

より初期の例を想定して,血清総ビリルビン値2.0,アルブミン値3.5,年齢65歳,プロトロンビン時間延長2.0秒,浮腫スコア0の患者の場合(条件2),リスクスコア4.73となり,1年後の生存確率は0.979となる。

これら2つの条件の患者の生存確率を他の年についても求めて,生存曲線を描くと図1のようになる。リスクスコア7.8未満で肝移植をしたほうが,それ以上の場合よりも予後が良好とされている³⁾。

表1 肝硬変症の予後指数と生存率の関係

予後指数	1年生存率	2年生存率
1	100	100
1.5	98	96
2	96	92
2.5	92	90
3	90	84
3.5	84	75
4	75	60
4.5	60	45
5	45	27
5.5	25	12
6	12	5
6.5	5	0
7	0	0
7.5	0	0
8	0	0

原発性硬化性胆管炎

原発性硬化性胆管炎(primary sclerosing cholangitis:PSC)のための改訂 Mayo リスクスコアはオリジナルから,肝生検の所見を含まない形に改訂されたものである⁴⁾(☆☆)。リスクスコアが1.0増加するごとに,死亡リスクは2.5倍に増加する。

リスクスコア $= 0.0295 \times \text{年齢(年)} + 0.5373 \times \text{血清総ビリルビン値(mg/dl)の自然対数} - 0.8389 \times \text{血清アルブミン値(g/dl)} + 0.5380 \times \text{AST(IU/l)の自然対数} + 1.2426 \times (\text{静脈瘤からの出血あり} 1.0, \text{なし} 0)$

リスクスコア0以下は低死亡リスク,0~2は中等度,2以上は高死亡リスクと判定される。

消化管出血のない肝硬変症

Abad-Lacruzら⁵⁾(☆☆)によって,Child-Pughスコアに腎機能,栄養,飲酒などの影響を取り入れ作成された多変量のモデルを紹介する。

予後指数 $= 0.302 \times \text{Child-Pughスコア} + 0.113 \times \text{BUN(mmol/l)} + 0.027 \times \text{MCV} - 0.025 * (\% \text{前}$

腕筋周径). $BUN(\text{mmol/l}) = BUN(\text{mg/dl}) \times 0.357$. %前腕筋周径は, 前腕周径(cm) $- 3.14 \times$ 上腕三頭筋部の皮膚をつまみあげた厚さ(cm)の健常者の中央値に対する%. したがって, その年齢の健常者の平均値の何%かという値なので推測値を用いることも容易である. MCVは平均赤血球容積, Child-Pughスコアについては成書を参照. 予後指数と生存率の関係を表1に示す.

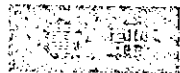
ここに紹介した以外にも多数のモデルが作成されている. The Medical Algorithms Project (<http://www.medal.org/index.html>)を参照されたい.

参考文献

- 1) 森實敏夫: 入門医療統計学, 東京図書, 2004
- 2) Dickson ER, et al: Prognosis in primary biliary cirrhosis; Model for decision making. *Hepatology* 10: 1-7, 1989
- 3) Kim WR, et al: Optimal timing of liver transplantation for primary biliary cirrhosis. *Hepatology* 28: 33-38, 1998
- 4) Wiesner RH, et al: Primary sclerosing cholangitis; Natural history, prognostic factors and survival analysis. *Hepatology* 10: 430-436, 1989
- 5) Abad-Lacruz A, et al: Routine tests of renal function, alcoholism, and nutrition improve the prognostic accuracy of Child-Pugh score in nonbleeding advanced cirrhotics. *Am J Gastroenterol* 88: 382-387, 1993



森實敏夫 著



入門 医療統計学 Evidence を見出すために



上野 文昭(大船中央病院特別顧問)

Sackett らが提唱し海外で定着した EBM は, ようやく最近わが国でも認知されるようになった. 適切にデザインされた臨床研究から得られる質の高いエビデンスは, EBM の重要な構成成分の一つである. 日常の診療や学会でもエビデンスという言葉を目にする機会が多くなり, 経験や意見のみに依存せず文献を調べ参照する姿勢が多く医師の間でも見られるようになったことは歓迎すべきである.

しかし書かれてあることを妄信する傾向が出てきたのは困った現象である. 世に出ている医学論文のすべてが正しいわけではなく, 適用上の限界は必ず存在する. 論文の内容, とくに研究デザインを吟味するためには基礎的な臨床疫学の知識が必要である. 中でも統計学は重要であるが, 臨床現場の医師にとってこれほど面倒なものはない. 簡単でわかりやすい教科書はほんの表面的な知識しか提供しないし, 今日の統計学を網羅した大作は読む気すら起らない.

このような時, 東京図書より「入門医療統計学—Evidence を見出すために」が上梓された. 著者の森實敏夫博士は元々肝臓病学と免疫学の優れた研究者であるが, 診療と研究をより正しい方向に導くために統計学を独学でマスターしたという. その熱意と才覚に敬意を表したい. 本書に目を通すと, 単に論文を正しく読むということを越え, エビデンスを創る側にも必要な知識も網羅されているようである. 内容は当然やや難解であるが, 例題や囲み記事などで興味をつなげようとする配慮が垣間見られる.

本書は入門書という枠を越えた膨大な統計学の情報を網羅した教科書である. しかしそのコンパクトなサイズとわかりやすい構成のため, 敷居の高さを感じさせない. 医学論文を正しく読みたい, よい臨床研究をしたいと願っている志の高い若手医師に真っ先に推薦したい一冊である.

(A5判 360頁 定価3,990円(本体3,800円+税5%))
(2004年 東京図書 刊)