

表：Evidence based Medicine の3要素とそのポイント

- 1) 根拠：最も妥当な、結果の明白な、状況に当てはまる
- 2) 価値観：患者の価値観・好み、社会の価値観・価値基準
- 3) 経験：医療者側の専門職としての経験・技能

ここで、この3要素をまとめておきましょう。

根拠：これが、医療情報・臨床研究結果です。これを上手に活用するためには、情報を広く求める能力（情報・文献検索能力）、その情報の信頼性を見極める能力（情報吟味能力）、その情報の現場での適用性や価値判断を行う能力（情報の適用能力・情報に基づいた判断能力）の3つが必要になります。このうち、情報の吟味能力に関しては少し臨床疫学に関する知識が必要です。

表：Evidence based Medicine で重要視される情報運用の3能力

- 情報・文献検索能力
- 情報吟味能力
- 情報に基づいた判断能力

残念ながら従来教育は「統計学」に関して数字を取り扱い、データを分析するというイメージを与えてきました。しかし、ここでの「臨床疫学」というのは、データの分析手法そのものに関してはほとんど知る必要性はなく、1) 研究デザインが一時的な思いこみや見落としが入りかねず信頼しにくいものであるかどうか、2) その研究結果がどういった意味を持つものであるか、2つの点を判断できることを目標にしています。決して有意差検定の手法を細かく知る必要はありません。

表：EBM で必要とされる臨床疫学の知識

- 研究デザインが一時的な思いこみや見落としが入りかねず信頼しにくいものであるかどうか：臨床デザインの種類の理解、様々なバイアスの種類とそれを避ける手法
- その研究結果がどういった意味を持つものであるか：治療効果の指標の理解、リスク比、危険減少率、絶対危険率減少、治療必要人数、などの指標の理解。検査・診断に関しては感度、特異度などの指標の理解。さらに信頼区間などの推計学的指標の理解

価値観：患者さんの価値観や好みに加えて、社会の価値基準からも医療における判断が無縁にはなれません。これを把握する能力も重要です。患者さんとのコミュニケーション技能、あらゆる価値観に対する共感する態度、これらを支える患者情報の把握、などが必要です。また、社会の価値基準からはずれないためには、その判断を共有する医療チームを形成することも重要になります。重症患者管理や救命治療、緩和ケアなど困難な判断が要求される分野であるほど、医療チームでのアプローチが強調されます。チームアプローチを行うと、同僚や他の医療職と判断を共有することで、判断のもととなる価値判断の基準が一時の思いこみや見落としを含んでいないかどうかお互いに確認することができます。

経験：医療側の経験や技能も判断に重要な意味を持ちます。手術や特別な処置の場合、施設の経験や技能によって判断が異なることはありえます。また、このような場合だけではなく治療に伴う不都合や様々な治療のオプションを知っておくことは、判断の広がりをもたせます。治療に伴う不都合や危険性、様々な代償、コストなどを医療側がわかまえていることも判断を行うときに重要なポイントになります。

この3つの要素はお互いの足りない部分を補うものです。例えば、明らかに生命予後の改善が期待される治療に関しては、患者さんの価値観とぶつかってもより強く説得ができるでしょう。逆に、もし有効性がはっきりしないものであれば患者さんの好みによって治療を行うか行わないかを決めて見よいかもかもしれません。ただ、「価値観」には患者さんのみならず「社会」の価値観を考慮する必要があります。また、すでに有効性が示された治療法であっても、複雑で訓練が必要で危険性も伴う場合には、その医療機関の判断としてあえて避けるという場合もあるでしょう。

さらに、医療や福祉、さらに社会そのものの資源（人員・人材、施設・設備、費用・予算など）の有無や質と量が判断に大きな影響を与えるため、必要な資源を第4の要素として別に挙げることもみられるようになっています。

<1-2> Evidence based Medicine の手法

はじめに

EBM とは 患者に対して、医療情報の妥当性・信頼性を十分ふまえた上で、確実に明確な臨床判断を行なうことを重要視する医療方法と定義されます。

これを実行するために必要とされるものは以下の3つです。

経験と専門知識

現状で利用可能な最も妥当な客観的根拠

これらに基づいた医療を可能にする環境

EBM は、実際の患者から必要な根拠、具体的には論文などを探し出す方法論から、そこから根拠を評価し、実際の患者に適用すべきかどうか、適用するならどう実際に行なうかを判断する方法まで含む、包括的で実際的な手法です。

Evidence-based Medicine の基本手順

- 今直面している疑問・問題を整理し、定式化する。→しばしばあいまいでわかりにくい課題を、お互いに共有しやすく答えを見つけやすいように整理する。
- この疑問に基づいて論文などの情報検索を行う。→問題の定式化の中で keyword は決まってくる。
- その情報を信頼してよいかどうか批判的に吟味する。→これも一定の手順に従って行う。
- その結果に基づいて、患者・状況にあった判断を行う。→ここでも一定のチェックポイントが示されている。
- その判断の結果を自己評価する。

役に立つ情報とは何か

研究方法が妥当で結果が信頼できること。

バイアスの入る余地が少なく、その結果が「真実」を反映していると考えられること。

- 研究デザインが信頼できる。
- 研究機関・研究者が信頼できる (?)。
- 載っている雑誌が信頼できる (??)。

結果が明らかであること

- 差が明らかである。
- 差が大きく、なおかつその差の大きさは偶然の産物ではない。
- 臨床上重要な差がない。
- 差がほとんどなく、なおかつたとえ偶然を考慮してもその差は重要なものとはなりえない。

自分の状況にあてはまり、自分もできる内容であること

- 自分の患者が研究の対象患者にふくまれる。
- 自分が自分の施設で行うことができる方法である。
- さらに必要なチェックポイント
- 自分の国や地方、医療体制、人種などの相違点を考慮しても、結果をくつがえすほど大きな影響はない。

さらに、「対象患者の価値観をふまえた判断ができるほど患者とコミュニケーションしているか」という項目が加えられることもあります。

〈II〉チェックポイント集

Users' guides to the medical literature: Evidence-Based Medicine Working Group (translated by T Fukuoka)
 CASP Japan ホームページ (<http://CASPjp.umin.ac.jp>) から入手可能

治療に関する論文

結果は信頼できるか

対象患者は治療群と対照群に無作為に割り付けられているか
 研究対象患者の全てが結果に反映されているか
 フォローアップ中に脱落症例はないか。あっても結果をくつがえすほどの影響はないか。

最初の割付のまま評価されているか (intention to treat analysis)

患者、医師、研究者がいずれも患者の治療内容を盲検化 (目隠し) されているか

最初の患者背景は両群で大きく異なっていないか

研究対象の治療以外は、両群の治療内容が大きく異なっていないか

結果はどのようなものか

結果はどれほど大きなものか (これは、効果の差を見る)

結果からの予測はどれほど正確か (これは、信頼区間の狭さを見る)

結果は自分のそして自分の患者の役に立つか

研究対象患者から見て、この結果は自分の患者に当てはめることができるか

重要な転帰が、全て検討されているか

その治療による益は、害やコストに見合ったものか

治療の論文での重要な指標

	転帰 (死亡など)	
	あり	なし
治療群	a	b
対照群	c	d

危険性 治療群: $Y=a/(a+b)$ 、対照群: $X=c/(c+d)$

相対危険度あるいはリスク比 (RR) $RR=Y/X$

相対危険度減少 (relative risk reduction: RRR) $RRR=1-RR$ or $RRR=[(X- Y)/X]$

絶対危険度減少 (absolute risk reduction: ARR) $ARR=X- Y$

治療必要人数 (number needed to treat: NNT) $1/ARR$

補足説明

死亡率 20%の疾患が治療によって死亡率 10%になった場合、RR は 1/2、RRR は 50%、ARR は 10%、NNT は 10 人となる。死亡率 4%の疾患が治療によって死亡率 2%になった場合は、RR は 1/2、RRR は 50%と先ほどと同様であるが、ARR は 2%、NNT は 50 人となる。

NNT とは、望ましい治療効果を呈する患者を 1 人得るために治療する必要がある人数を指す。治療によっては、時間の要因を考慮する必要がある。たとえば治療といっても急性期の治療ではなく、3 年間の治療継続が必要な場合には、3 年とする。

文献検索の場合の keyword

Clinical trial (publication type)

Randomized controlled trial (publication type)

診断検査に関する論文(Diagnostic test)

結果は信頼できるか

標準的な診断を確定する方法(gold standard)と、その結果を知らされないまま別々に検査され比較しているか
 研究の対象となった患者群は実際に検査の対象となるような患者群であったか

研究の対象となった検査は、標準的な検査結果にかかわらず全例に行われ評価されているか

検査方法は明確に記載されて実行・再現可能であるか

結果はどのようなものか

感度、特異度、尤度比はどうか

結果は自分のそして自分の患者の役に立つか

その検査は自分の施設、状況で行なえるか

結果は自分の患者にあてはまりそうか

その検査は自分の医療行為を変えうるものか

その検査によって患者に利益があるか

診断の論文の中での重要な指標

	疾患あり	疾患なし
検査陽性	a	b
検査陰性	c	d

感度 $a/(a+c)$ 、特異度 $d/(b+d)$

検査の陽性的中率 (positive predictive value: PPV) $a/(a+b)$

検査の陰性的中率 (negative predictive value: NPV) $d/(c+d)$

陽性の尤度比 (likelihood ratio of positive test: LR(+))

感度 / (1-特異度) or $(a/(a+c))/(b/(b+d))$

陰性の尤度比 (likelihood ratio of negative test: LR(-))

1-感度/特異度 or $(c/(a+c))/(d/(b+d))$

尤度比の目安

LR が > 10 あるいは < 0.1

効果大

LR が 5-10 あるいは 0.1-0.2

効果中

LR が 2-5 あるいは 0.2-0.5

効果小

LR が < 2 あるいは > 0.5

効果僅か

LR が 1

効果なし

LR が大きければ大きいほど、確定診断に役立つ。逆に小さいほど除外診断に役立つ。

文献検索の場合の keyword

Sensitivity and specificity --Explode

Diagnosis &(px)

予後に関する論文

結果は信頼できるか

対象となった患者はその疾患の同等の病期（重症度、進行度）から得られているか

追跡は十分長く十分完全に行なわれたか

転帰は客観的でバイアスを避けたものが用いられたか（死亡や詳細に定義された合併症）

重要な予後決定因子に関して補正を行ない検討されているか

結果はどのようなものか

経過に伴う転帰はどのようなものか（予後の差を見る）

予後に関する予測はどの程度正確か（その差の信頼区間を見る）

結果は自分のそして自分の患者の役に立つか

対象患者は自分の患者と似ているか

その結果が治療の選択に関わるか

その結果を語ることが患者の役に立つのか

予後の論文の中での重要な指標

予後の記述は、5年生存率××％というように絶対的な値を示す場合、危険因子などの有無で比較して相対的に示す場合、実際の生存曲線を示し時間経過を通じて記述する場合がある。実際の生存曲線を示し、時間経過に伴う転帰を示した方が情報量は多い。

予後を決定する因子としては、患者の身体的な問題（性別、年齢、人種など）、病気の進行度（癌の stage 分類など）、合併症（高血圧や糖尿病、検査の異常値など）が検討される。予後を予測するために患者を集めた場合には、患者を2群に分けて検討する。まず、derivation sample（誘導群）で危険因子などを検討し、それとは別の validation sample（確認群）でその予測がどれほど確からしいかを検討する。

文献検索の場合の keyword

Cohort studies---Explode

Prognosis (text word)

副作用に関する論文

結果は信頼できるか

比較する患者群は、その副作用の原因と考えられる要因以外に大きな違いはないと思えるか

副作用の原因に関しても、副作用に関する転帰予後に関して両群の調査方法、評価方法に大きな差がないか

追跡は十分長く十分完全に行なわれたか

時間的前後関係は正しいか（結果の前に原因があったか）

量反応関係がなりたつたか（原因が大きいほど結果も大きかったか）

結果はどのようなものか

要因と結果との間に強い関連があるか（相対危険度やオッズ比の大きさを見る）

危険性はどれほど正確に予測されるか（その値の信頼区間の狭さを見る）

結果は自分のそして自分の患者の役に立つか

その結果は自分の患者にあてはまりそうか

自分の患者の危険はどの程度のものか（患者の背景や病状、その他の要因などから予測する）

その要因を排除しようとするべきか（代替案があるか、排除しても大きな問題はなさそうか）

副作用の論文の中での重要な指標

副作用を示す論文は副作用が疑われて検討される場合が多いために、通常 case-control study（症例対照研究）のデザインとなる。前向き cohort study のほうが信頼度は高くなる。

	副作用（死亡など）	
	あり	なし
要因あり	a	b
要因なし	c	d

RCTや prospective cohort study では相対危険度（relative risk: RR）が求められる。

$$RR = (a/(a+b)) / (c/(c+d))$$

RR が 3.0 ということは、その要因があった場合には、要因がない場合に比べて副作用発現の危険率が 3 倍になることを意味する。

しかし、case-control study では RR は求められずオッズ比（Odds ratio: OR）を求める。

ポイント：case-control study では、副作用がわかってから症例を集める。このため副作用があった群（case）とその対照である副作用がなかった群（control）の数の関係は研究のデザインによって変わる。したがって、副作用の発生率自体をこの表から $a/(a+b)$ とか $c/(c+d)$ といった式で求めようとする、1 case あたり 1 control 集めた場合と、2 control 用意した場合とでは異なった値となる。したがって、case-control study でこの 2 分割表を水平方向に比べるのはそのこと自体誤りであり、垂直方向に副作用のあった患者のどれくらい要因があり、要因がない場合はどれくらいかということから、副作用との関連を検討することになる。

$$OR \text{ (オッズ比)} = (a/c) / (b/d) = (a*d) / (b*c)$$

文献検索の場合の keyword

risk (text word)

Cohort studies---Explode

体系的レビューメタアナリシス

結果は信頼できるか

そのレビューの課題の焦点は十分絞られたものか
レビューの対象とする論文の基準は適切か（バイアスの入りにくい妥当性の高い研究を選んだか）
重要な研究が見逃されていないようにされているか（複数のデータベースを検索したか、言語の制限は設けたか、検索論文の reference も対象にしたか、専門家や企業などにも問い合わせたか）
対象となった論文の信頼度に関する批判的吟味が行なわれているか
それぞれの研究の評価は一定か（検討された研究で結果は一定方向であるか）
それぞれの研究の結果は同等か（研究結果の信頼区間は重なっているか）

結果はどのようなものか

レビューの全体の結果は何か（重要な臨床上の転帰を見極める）
その結果はどれほど正確と考えられるか（信頼区間を見る）
結果は自分のそして自分の患者の役に立つか
その結果は自分の患者にあてはまるか
重要な転帰はすべての考慮されているか
その益は、害やコストに見合ったものか

体系的レビュー/メタアナリシスでの重要な指標

レビューされた研究をまとめる前に、それぞれの研究が同質なものであるかを検討する必要がある。それぞれの研究の結果が偶然から予想される範囲（信頼区間）を越えてばらついていないかどうか検討する。通常は統計学的手法を用いて、均質性（homogeneity）などを検討する。

文献検索の場合の keyword

meta-analysis (publication type)
review (publication type)

経済的分析に関する論文(Economic evaluation)

結果は信頼できるか

研究で行われた経済的（金銭的）比較は、医療行為全体を対象としているか
コストと転帰・結果の計算方法・評価方法は適切か
分析の中で不明確な部分に関しては、適切な幅が設定されていたか
コストと転帰・結果は、実際にその治療の対象となる患者を想定して計算されているか
結果はどのようなものか
それぞれの対策に必要なコストと転帰・結果はどれほどであった。
必要なコストと転帰・結果はサブグループによって異なるか。不明確な部分による結果への影響はどれほどか（感度分析など）。
結果は自分のそして自分の患者の役に立つか
治療による益は害やコストに見合ったものか
自分の患者でも同等の健康上の転帰が期待できるか
コストも同等であると思われるか

経済分析での重要な指標

経済分析とは単に金銭のみ扱うものではない。選択肢のコストと結果の比較を行うものである。

Full economic comparisonとは、いくつかの医療行為・政策のコストと結果を比較したものであり、以下の3種類がある。

Cost-benefit analysisとは利益とコストを金銭の尺度を用いて示し比較したものである（利益もお金に換算する）。

Cost-effective analysisとはその施策のコストを臨床上の尺度によって比較したものである（たとえば、生存期間の延長など）

Cost-utility analysisとはその施策のコストと社会的な尺度を用いて比較したものである（たとえば、quality adjusted life year: QALY など）

Direct Health Costとは、その医療施策に実際に費やされる費用・リソースから求められる

Indirect Health Costとは、その医療施策によるその他のリソースのことである（たとえば、収入の減少など）

Discountingとは、将来のコストを現在の価値として比較するために行われる（10年後の10000円と、今の10000円とは同じ価値ではない）。

文献検索の場合の keyword

economic analysis (textword)
cost-benefit analysis (MeSH)

治療ガイドライン (Practice guidelines)

結果は信頼できるか

重要な治療の選択肢とその結果が全ての検討されているか
根拠 (evidence) を集め、選択し、まとめあげた手順は明確で実際的なものであったか

異なる転帰の比較の手順も明確で実際的なものであったか

最近の重要な知見も含まれているか

そのガイドラインはピアレビュー (同僚の評価) と評価を受けたものか

ガイドラインの推奨はどのようなものか

実行可能で重要な指示・推奨がなされているか

その指示はどれほど強く推奨されるか

根拠や価値判断の不明確な部分がガイドラインの推奨を変える可能性があるか

ガイドラインの推奨は自分のそして自分の患者の役に立つか

ガイドラインの目的は自分のものと一致しているか

その推奨は自分の患者にあてはめることができるか

治療ガイドラインでの重要な指標

治療ガイドラインとは、医師と患者が行う、ある医療状況下での決断が適切なものになるように補助することを目的に、体系的にまとめられた声明であり、以下の要素を有する。

1) 根拠の包括的なレビュー、2) 価値判断 (重みづけ) のための専門家の意見、3) 正式な承認と支持、4) 実際の医療行為に影響を与えようとする意図

価値とは、異なる臨床上的転帰・結果に関する個々人の好み・選択のことである。

患者の好み・選択とは、ある健康状態に対する患者の主観的な満足や苦痛、望ましさなどのことを指す。

ガイドラインの推奨のグレード

A1: 複数の RCT、結果の不均一性はない、信頼区間は NNT の閾値をクリアしている

A2: 複数の RCT、結果の不均一性はない、信頼区間は NNT の閾値をまたいでいる。

B1: 複数の RCT、結果は不均一、信頼区間は NNT の閾値をクリアしている

B2: 複数の RCT、結果は不均一、信頼区間は NNT の閾値をまたいでいる

C1: 複数の観察研究、信頼区間は NNT の閾値をクリアしている

C2: 複数の観察研究、信頼区間は NNT の閾値をまたいでいる

文献検索の場合の keyword

practice guideline (publication type)

決断分析 (Decision Analysis)

結果は信頼できるか

重要な手段と転帰が全て考慮されたか

それぞれの転帰の確率を求めた根拠は、明確で実際的な手段によってまとめられているか

転帰の評価 (utility) は信頼できる情報から、明確で実際的な手段によって得られたものであったか

それぞれの証拠の不確かさの影響がどれほどであるかについて検討されているか

結果はどのようなものか

結果全体として、ある一つの手段が臨床上重要な利益を患者に与えることが示されたか もしそうでなければ、結局運任せということか

この分析に用いられた根拠の強さはどれほどと考えられるか
根拠の不確かさが結果を覆しうるか

結果は自分のそして自分の患者の役に立つか

それぞれの確率は、自分の患者の臨床像に合致するか

転帰の評価 (utility) は自分の患者の価値観が反映されているか

決断分析とは、不明確な状況下での決断を分析するために行われる、体系的で明確な量的な解析である。体系的な常識の結合とも言える

〈III〉情報のチェックポイント

ほすびたるらいぶらりあん 2000 年 vol25

Page 86-96.

具体的な情報のチェックポイントを考えてみましょう。EBM の手順などで示される批判的吟味での情報のチェックポイントの3つの要素を表5に示します。具体的な手法に関しては、他にまとめられています¹²³。

ここでは、論文の種類ごとに最も重要なポイントを2つずつまとめて解説します。このチェックポイントの一覧表を表6に示します。この重要なチェックポイントは、その論文が読むに値するかどうかの目安ともされます。実際のチェックポイントはもっと項目数も多いのですが、論文の価値の目安を身につけるために、とりあえずこれらのポイントを理解しておきましょう。

表5 情報のチェックポイントの3要素

1) <u>情報は信頼できるかどうか</u> : 情報のもととなった研究方法は、様々な思いこみや見落としが避けられるよう、注意深く計画されたものであったか。また、実際に対象とした患者やその追跡なども、結果をくつがえすような問題をはらんでいなかったか。
2) <u>情報の意味することは何か</u> : どの程度臨床での意味がありそうか (治療効果や診断の感度特異度、予後判定など)。また、それはどれくらいあいまいさを残していると判断されるか (推計学的信頼区間など)
3) <u>その情報は役に立ちそうか</u> : 自分の状況に当てはまるか。実行可能か。患者や家族もそれを望んでいるか。社会もそれを容認しているか。

表6 論文のチェックポイント: ポイント中のポイント

<u>すべての論文に共通するチェックポイント</u>
● 論文の課題は明確に定義され十分焦点が絞られたものであったか
<u>治療の論文のチェックポイント</u>
● 患者の割付は無作為か (ランダム割付に基づく臨床試験であったか)
● 研究対象患者の全てが漏れなく結果に反映されたか (追跡できなかった症例は少ないか <u>Intention to treat analysis</u> か)
<u>診断の論文のチェックポイント</u>
● 研究対象の検査結果は、確定診断の結果と別々に判定され、お互いの結果を知らされずに検討されたか
● 研究は、実際にその検査の対象となると思われる患者群において検討されていたか

¹ 川村孝. 批判的吟味のポイントー治療、予防について. 臨床と薬物治療. 1999; 18: 967-971.

² 川村孝. 批判的吟味のポイントー診断、予後、レビューについて. 臨床と薬物治療. 1999; 18: 972-976.

³ Greenhalgh T (今井二郎, 渡邊聡子訳). EBM がわかる. 金芳堂. 1999.

<u>病因/副作用の論文のチェックポイント</u>
● 比較検討したグループは、原因と思われる要因以外は同等であったか
● 原因と思われる要因もその転帰 (副作用) も両群で同じように検討したか
<u>予後の論文のチェックポイント</u>
● 対象患者の診断基準は明記され、同じような病期 (進行度、重症度) の患者群を対象に検討されたか
● フォローアップは十分長く完全であったか
<u>体系的レビュー・メタアナリシスのチェックポイント</u>
● レビューで検討した課題は、焦点の絞られたものであったか
● レビューに用いた論文の選択基準は適切であったか

A) すべての論文に共通するチェックポイント

- 論文の課題は明確に定義され十分焦点が絞られたものであったか

このチェックポイントはすべての情報・論文に共通です。臨床論文では対象患者に対して何らかの治療法や診断検査などを行ったかどうか、あるいは時間的経過や危険因子の有無などによって、臨床経過や診断、転帰などに違いがなかったかを検討しています。これらの要素が明確に記載されていなければ、評価も解釈も適用もしにくいものになります。どのような患者であったのか、どのようなことを行ったのか、何で評価したのかを把握できるなければなりません。これは、「目的」または「方法」に記載されています。

B) 治療の論文のチェックポイント

- 4) 患者の割付は無作為か (ランダム化比較試験であったか)

このランダム化の重要性は前に述べました。割付方法については「方法」の部分に記載されています。多くのランダム化比較試験の場合、「ランダム化した」という記載だけではなく、どのような手順でランダム化したかについても記述されています。

- 5) 研究対象患者の全てが漏れなく結果に反映されたか (追跡できなかった症例は少ないか Intention to treat analysis か)

研究途中に追跡できなくなり予後がわからない症例があると結果の妥当性を脅かします。最低限の追跡率の目安は80%以上とされています。しかし、このため、検討された評価基準の頻度が少ないほど、より厳しい判断が必要になります。たとえば、死亡率が1%程度の疾患で死亡率を検討した研究の場合、5%の追跡不能症例がいるだけでその症例の経過・転帰によっては研究結果が覆ってしまいます。この点も検討して読みとる必要があります。追跡不能例については多くの場合「結果」の冒頭に記載されています。

薬剤の臨床研究であった場合、たとえ内服を中止した症例であっても結果は評価する必要があります。また、内服を中止した症例であっても、それは「その薬剤を内服した群」の一人として解析に加える必要があります。これを、intention to treat analysis (治療する意図に基づく解析) といいます。実際に治療したかした

いかではなく、最初にランダム化して割り付けたそのグループのまま解析する方法です。こちらの方が、バイアスが入りにくく、治療効果の判定が厳しくなり、現実的な評価にないます（現実には薬を処方しても薬を飲めなかったり飲まなかったりする患者はいます。ですので、このような要因も含めて検討するのが実際のです。）。最近の論文では「方法」の部分に「解析はintention to treat analysis によった」と明記されることが多くなってきました。明記していなければ、患者背景の表と結果の表とを比較して患者数に違いが見られれば、intention to treat analysis ではないという風に確認する方法もあります。

C) 診断の論文

- 研究対象の検査結果は、確定診断の結果と別々に判定され、お互いの結果を知らされずに検討されたか

評価したい検査結果と確定診断の結果がお互いに目隠しされていなければ正確な検査の有効性はわかりません。検査結果を知った上で確定診断を行うと検査結果と診断の関連が強まってしまう。また逆に、確定診断を知ってから検査結果を判断する場合にも結果に影響がでる可能性があります。これを避けるためには、検討する検査の判定と、確定診断を行う作業は、別々の医者がお互いの検査結果・診断結果を知らないままで行う必要があります。これが重要なポイントになります。治療におけるランダムかと同じくらい重要なポイントです。これは、「方法」の検査方法と確定診断の方法などの前後に記載されています。

- 研究は、実際にその検査の対象となると思われる患者群において検討されていたか

実際の臨床の現場において、検査を行う患者は何らかの疑いを持った患者群になります。たとえば、急性心筋梗塞に関する新しい診断方法が開発され検討された場合、急性心筋梗塞患者と正常人ボランティアとで比較した研究はあまり意味がありません。この両者を区別するために検査をすることは考えにくいからでしょう。これを確認するには、「方法」の対象患者の定義のところをチェックすればよいことになります。

D) 病因/副作用の論文

- 比較検討したグループは、原因と思われる要因以外は同等であったか

何らかの危険因子や原因と思われる要因が実際に病気や特殊な臨床経過の原因になっているかどうかを検討する研究では、その要因のある群とない群とで比較することになります。このとき、要因の有無をランダム化して決めることは必ずしもできません。このため、原因以外の要因をそろえる作業が必要です。喫煙の影響を調べる研究であれば、対象とした喫煙者と背景の同じ非喫煙者を探して対照群を作る、対象患者の年齢や背景を制限し大きくばらつかないようにして検討する、大きく異なる因子に関して統計学的に補正を行うて検討する、などの工夫がなされます。これらの工夫によって検討された要因以外の背景をがそろえられて

いるかを見極めましょう。これは、「方法」のどのよう

- 原因と思われる要因もその転帰（副作用）も両群で同じように検討したか

原因と転帰を比較する2群がうまく定義できたとしても、要因や転帰の検討方法に大きな差が生じては結果に大きな影響を与えます。例えば、たばこ胃潰瘍の関連を調べる研究を行ったとして、喫煙者だけが定期的に検診を受けさせていたり、あるいは本人任せにしたために非喫煙者が熱心に検診を受けたりすると、胃潰瘍の診断の多少がたばこによるものなのか検診を受けたか受けないかによるものなのかがわからなくなってしまいます。逆に、膵癌とアルコールとの因果関係を調べるために、膵癌の人とそれと似た膵癌でない人とのアルコール摂取量を検討した場合も、一方は入院中の聞き取り調査でもう一方は単に郵送による調査では答えに大きな差が生じる可能性があります。このように、同じようにそれぞれの要因や転帰を同じように調査評価していたかどうかを検討する必要があります。これは、「方法」で要因や転帰の定義や調査・検討方法を見ることになります。

E) 予後の論文

- 対象患者の診断基準は明記され、同じような病期（進行度、重症度）の患者群を対象に検討されたか

対象患者の定義もあいまいでその病気の進行度なども検討しないまま予後の検討を行っているのは、どのように読みとっていいのかがわかりません。肺癌では、腫瘍の大きさやリンパ腺転移の有無、他の臓器への転移の有無などによって病期が定義されます。さらに、扁平上皮癌が腺癌、小細胞癌などの病理診断によって大きく異なります。このような要因をふまえて検討された予後調査である必要があります。これは、「方法」の部分で対象患者の定義の部分に記載されています。

- フォローアップは十分長く完全であったか

予後調査に関しては十分な経過を検討する必要があります。また、治療の論文と同じように追跡不能症例が多くなると臨床経過上のあいまいさが大きくなります。どれほどの追跡率が十分と考えるかは、治療の論文と同じでよいでしょう。80%が目安となり、その経過中で検討した死亡率などが小さいほど、わずかな追跡不能症例が大きな問題になります。追跡期間かどうかは、その疾患の経過や検討した評価基準によって異なります。たとえば、癌などの疾患の予後を数ヶ月で検討した研究では不十分と思えるかもしれません。一方、咽喉炎の予後を1年以上検討しては不必要と思われるかもしれません。しかし、同じ咽喉炎に関する検討でもその後の呼吸機能にどのような影響を与えるか、といった研究テーマでは当然長期の追跡が必要になります。通常、追跡期間は「方法」の部分に、追跡率は「結果」のところに記載されています。

F) 体系的レビュー・メタアナリシス

- レビューで検討した課題は、焦点の絞られたものであったか

体系的レビューとは、ある課題を検討した過去の論

文をすべて集めて検討するものです。検討した論文の治療効果などをすべて足し合わせて検討する手法がメタアナリシスです。この手法が妥当であるためには、課題が明確に定義されている必要があります。たとえば、「感染症に抗生物質は効くか」というのでは、どんな感染症なのか、どんな抗生物質なのか、効くというのはどのようなことを指すのか、全くわかりません。

「ICUに入室した重症患者に対して、経静脈的に抗生物質を投与すると、ICU死亡率が低下するか」というと定義は明確になり焦点が絞られたことになります。レビューの課題は、通常論題に示されていますし、「方法」や「目的」にも記載されています。

● レビューに用いた論文の選択基準は適切であったか

レビューのために探した論文の選択基準が不適切であると、足し合わせることは危険になります。まず、最初の課題に合った論文が取り扱われたかを確認しましょう。次に、バイアスの入りにくい質の高い研究が検討されたかをチェックします。たくさんの論文を検討する利点の一つに、結果を足し合わせると症例数が大きくなり偶然によるばらつきの影響は小さくなる点があります。しかし、一時的な思いこみや見落としの影響は足し合わせてもなくなりません。むしろ結果に大きな影響を与えてしまいます。実際に、ランダム化されていない比較試験の結果を足し合わせても、真の結果は予測できないことが示されています⁴。治療に関するレビューであればランダム化比較試験を対象としている必要があります。これらは、いずれも「方法」の論文の選択基準や検索方法などを見ると判断できません。

4) さらに批判的吟味を学びたい方に

ここに挙げたチェックポイントは、最も重要なポイントのみです。他にもいくつかのチェックポイントがあります。チェックポイントを明確にして論文を読む利点としては、一時的な見落としや思いこみが避けられる、真に重要なポイントに焦点が当てられるため全体の労力は軽減する、等があります。批判的吟味をするためには論文のチェックポイントを知っておく必要があります。

さらに、批判的吟味を学ぶためには、それぞれのチェックポイントに従って論文を読むトレーニングを積むことが近道です。最近国内でもこのトレーニングを目的としたワークショップが盛んに開かれるようになってきました。このようなワークショップに参加されることが、重要な機会になると思われます。また、批判的吟味は自らの行動の一部として身につけることが重要です。日頃から批判的吟味を一人で行い続けることは困難ですが、職場の中でチームとして取り組むことで一人一人の負担は軽減されるでしょう。また、同じ職場で論文を読み合わせる抄読会を開くことが有効と思われます。この中で、お互いの吟味内容をチェ

ックし合うことで、思いこみや見落としに気づくことができます。情報を吟味する作業が日常業務の中で行えるようになれば、批判的吟味はかなり容易なものになります。実際の論文の批判的吟味を紹介したものもありますので、これらも参考になるでしょう⁵。

批判的吟味の技術を身につけると、文献検索結果を提供するときに、すべてを提示するのではなく情報を質の高いものだけに限ることができます。実際に情報を利用するものにとっては、情報が羅列的な数百件の論文のリストではなく、その中で質の高いものに絞ったリストの方が利用価値は高いはずですが、確かに、病態生理学的知識や臨床経験という点では病院図書館員と医療スタッフと大きな差があるかもしれません。しかし、情報の吟味は作業として定型化できる部分も多く、決して専門知識が必須ではありません。また、吟味を通して臨床での情報の活用方法などについて理解が深まると期待されます。最近、臨床のガイドラインやクリティカルパスなどが作られる機会が増えてきました。これらの作成過程には過去の論文のリストアップとその評価が不可欠です。この中で、網羅的な検索と定型化された吟味作業へのニーズが高くなっています。

⁴ Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomized clinical trials. *BMJ* 1998; 317: 1185-1190.

⁵ 福岡敏雄, 武澤純. 医療行為の決定法: EBM とはなにか. *Jpn J Endourol ESWL*. 1999; 12: 21-27.

⁶ 福岡敏雄編. 特集 critical appraisal skills 文献の批判的吟味 臨床と薬物治療. 1999; 18: 963-1029.

<IV-1> 患者への適用のポイント

海外のデータの利用に関する Users' Guide

(Dans AL ら JAMA 1998; 279: 545-549)

海外の研究結果を自国にあてはめるときには、以下の3つのポイントに注意する必要があります。

1) biologic

2) social and economic

3) epidemiologic

1) biologic

- その論文の結果を弱める方向に働く、自国の患者の病態生理学的な違いがあるか。

たとえば、同じ心不全、呼吸不全といってもその原因疾患は地域によって異なる。ある特定疾患を対象にした治療法を、ある疾患群に行なう場合には、そのサブグループの内容が大きく異なれば、全体としての効果に大きな影響を与える。

- その結果を弱める方向に働く、患者の違いがあるか。

たとえば、薬剤を代謝する酵素(N-acetyltransferase)の活性が日本人を含むアジア人では高い人が多く、ある種の薬剤は増量しなければ十分な効果が得られない。

2) social and economic

- その結果を弱めるように働く、患者のコンプライアンスの違いがあるか。

たとえば、経済状態や医療保険制度によっては継続的な治療が著しく困難な場合がある。

1) その結果を弱めるように働く、医療者・医療供給システムの違いはあるか。

その研究で必要とされる検査や治療が行われていなかったり、必要な薬剤が認可されていなかったりする場合がある。

3) epidemiologic

- 治療の効果や副作用を大きく変えるような患者背景があるか。

たとえば、ある種の診断器具の普及率が地域ごとに大きく異なる場合がある。また、患者の死亡原因や合併症も地域・病院によって異なる場合もある。

フィリピンのデータでは心筋梗塞の死亡原因の3分の1は肺炎・敗血症であり、このような状況では抗生剤の予防投与が血栓溶解療法より有効かもしれない(上記の論文での記載)。

- その研究結果に影響を与えるような、対照群(非治療群)に死亡率や合併症発生率の違いはないか。

たとえば、もともと日本人には心筋梗塞などの虚血性心疾患の頻度は低い(欧米の10分の1かそれ以下)。また、急性心筋梗塞の死亡率も欧米に比べて低いことが知られている。このような場合、虚血性心疾患の予防的治療の有効性は欧米での研究よりも少ない可能性がある。逆に、脳血管疾患の予防効果は大きく出る可能性がある。

<IV-2> EBM の落とし穴と日本での課題

EBM の落とし穴

(Feinstein AR ら Am J Med 1997; 103: 529-535)

● 研究で評価される情報

soft data の評価は？：「あいまいな訴え」「スコア化できない症状」これらをバイアスの温床であるからといって、無視して良いのだろうか。

● 研究で評価される患者

研究で評価されるような典型的・平均的な患者は必ずしもすべてではない。

● 研究の結果の当てはめ

統計学的有意水準に達しない治療は、行なうべきなのか、行なうべきでないのか。

たとえ有用性が証明されていても、何らかの理由で実行できない治療、診断手技もある。

● RCT の実行可能性

臨床の現場に必要な情報があっても、RCT が実行可能であるとは限らない。

● EBM そのものの抱える問題

EBM そのものが権威化し硬直化する可能性はないか。その中で、利用される情報に偏りが生じればバイアスが生まれる。

医療費抑制という強い社会的要請によって、今後さまざまな治療法を否定することを目的にした動きと結び付き、根拠が不十分だからというだけで有効かもしれない未評価な介入を早々に抹殺するように使われる危険性がある。

日本での EBM の役割

日本で EBM を行う障害

1) EBM に必要な情報の利用が困難

Medline などのデータベースが一般的に利用可能でない。

多くの情報が外国語である。

日本語のデータベースが充実していない。

日本語の総説や論文紹介は二次情報として不適切なものが多い。

2) EBM を実践する動機付けに乏しい

病態生理中心の教育、臨床研究より病因病態研究の重視

出来高払いに基づく医療保険制度と Evidence とのねじれ現象

結果に重点を置いた評価が一般的で、判断そのものの妥当性は軽視されがち

日本で EBM を身につける意義

1) 医師個人として

医療情報の整理：あふれる医療情報の中で何を捨ててよいかの基準が明らかになる

治療方針決定手順とその結果予測の明確化：インフォームドコンセントへの応用

今後の医療保険制度改革に伴う変化への対応

2) 医学医療の専門集団として

医療チームとしての明確な合意形成

専門集団として社会に対する説明責任

社会からの「質」の評価に耐えられる医療内容の確保

3) 社会として

医療情報の開示：とくに医療費の適正な配分に必要な情報の確保

医療内容の標準化とその評価基準の明確化

社会全体での政策として、医学医療に対する意義付けの合意形成

〈付録 1〉 ガルシニアダイエットのエビデンス

「助産婦雑誌」2000年9月号 巻頭言

「ねえ、ガルシニアダイエットって知ってる？」
「えっ、あなたダイエットしてるの？」
「ちがうわよ。ほら、時々女性雑誌に載っているでしょ。インド原産の薬草よ。これにやせる効果があるかどうか調べた研究を見つけたのよ。これが面白いんだから。」
「ほんと！？ で、やせるの？」
「結論を急がないの。あせって結論だけ知ったって、ニセ情報だったら困るでしょ。信用できるかどうかちゃんと確かめないと。ほら、これがその研究なのよ。」
「げっ。これ医学論文じゃないの。」
「だからすごいんじゃない。雑誌広告やテレビ番組と全然違うのよ。真剣に調べてるんだから。もとは英語の論文なんだけど日本語訳だから読みやすいでしょ。」
「ちょっと見せてね。なにに、地方新聞に広告を載せて、それに応募してきた人に協力してもらったわけね。対象になるかどうかは BMI という肥満の指標を使ったのね。そして、ガルシニアの有効成分を含んだ錠剤を飲む人と、プラセボというその成分を含まないニセものの錠剤を飲む群に分けて比較してるんだ。しかも、どっちの群にするかは、乱数に従って決めたのね。」
「これがランダム化よ。単に飲む飲まないで比べたら、ダイエットに並々ならぬ情熱を持った人ばかり飲むかもしれないでしょ。それじゃあ、ガルシニアの効果なのか、その人の熱意のせいなのかわかんないじゃない。」
「なるほどね。参加者は2週間ごとに受診して、12週間追跡したのね。」
「ちょっと短いかもしれないわね。1ヶ月でどうだったか判断するよりはまだけど、1年くらい追跡してくれても良かったかもしれないわ。ねえ、その評価方法のところをみてよ。」
「体重はデジタル体重計でクリニックで測定し、体脂肪量はレントゲンのスキャナーを用いて測定してるわね。単に写真を比べたり、自己申告じ

やないんだ。これじゃあ、うそもつけないわ。」
「しかも、誰がガルシニアを飲んでいるかわからないから、有利な手心も加えられないのよ。」
「盲検化って言うやつね。誰が何を飲んでいるのか、本人にもそれを見ている人にもわからないようにしておくんでしょ。確かに間違いは起こりにくいわね。何でどうやって測ったか詳しく書いてあるわね。くどいくらい。」
「結論だけ知りたいのだったらむだに見えるかもしれないけど、結論が信用できるかどうか知りたいのなら、ここまで書いた方が親切でしょ。」
「確かにそうね。やったことがいいかげんだったら、気が引けてここまで書けないわよ。ちゃんと書いているだけで信頼できそうな気がするわ。」
「それにほら、この図がいいでしょ。研究の対象となった人の結果が一目瞭然よ。180人対象となりそうな人がいて、そのうち45人は医学上の問題や予定の都合がつかなくて止めていて、135人が対象になっているでしょ。」
「結局、研究の対象になった人がどうなったかが全部示されているのね。これを見ると、135人のうち69人がプラセボを飲んで、66人がガルシニアを飲んだわけね。でも、そのうちの何人かはクリニックに来なくなったりして途中で止めているわよ。」
「そうなのよ。そこがこの研究の弱いところなの。結局12週の最後まで研究を続けられたのはそれぞれ60%くらいずつなのよ。命に関わるような病気じゃないし、しかも基本的に元気な人に外来で行っているわけだから、途中で止めたくなるのも分からないではないんだけど。」
「でも、ダイエットなんだから入院してやるのも変だし現実的ではないわよ。これくらいが限界なのかしら。」
「そうね。」
「で、結果は...え！！」
「ねっ、面白いでしょ。」

参考：Heymsfield SB ら：やせ薬としての *Garcinia cambogia* (ヒドロキシクエン酸) JAMA 日本語版 1999年10月号104-110頁

<付録 2>検査結果を読みとる・医学教育の現場で一

「臨床と検査」2002年3月号 p305

シナリオ

君の友人には2年前からつき合っている彼女がいた。2人はそばから見ていてうらやましいほど仲が良く、友人は彼女と将来結婚してもよいと言いはじめていた。

その友人が、突然連絡を取ってきてあって欲しいという。会ってみると、彼は途方にくれていた。話を聞くと、彼女は妊娠し、しかも検査結果の結果、彼女の HIV の血液検査が陽性であることが判明したという。

「僕の子だ。でもこのままだと彼女はどうか？ 子供はどうか？ 僕はどうすればいい？」と知っている。

彼女は悲観にくれながらも、感染症専門医と相談を始めているという。

今の彼と彼女がどのような課題に直面し、どのような情報が必要なのか考えてみよう。

このシナリオは名古屋大学での選択授業の中で用いているものである。これを使った学生との議論の詳細は別に執筆する機会を頂いた⁷。できれば、これに目を通してから以下を読み進めていただきたい。

このシナリオから、検査結果の解釈について学生から疑問が提起されることがある。

「HIV の検査が陽性だからといって、HIV に感染したと言えるのでしょうか？」

このすばらしい疑問に対して、わたしはこう答えるようにしている。

「ある教科書では、Enzyme Immunoassay による HIV 検査の感度も特異度も 99% かそれ以上と記載されている⁸。彼女が陽性であった検査が、感度 99.5%、特異度 99.5% であったとして、この彼女が HIV 感染者である確率はどれくらいだろうか？」

厳しい事実が学生たちは当惑する。そして、「彼女の確率が 95% 以上であると思う人？」という質問に、ほとんどの学生が手を挙げる。

「では、ここで感度と特異度についてもう一度考えてみよう。だれか、2×2表を使って説明

してくれないかな。」と話を向けると、授業に協力してくれる学生が出てくる。「少し、不確かなんですが、…」といいながら、以下のような図を書く。

	疾患あり	疾患なし
検査陽性	真の陽性	偽陽性
検査陰性	偽陰性	真の陰性

感度=真の陽性/(真の陽性+偽陰性)

特異度=真の陰性/(真の陰性+偽陽性)

「すばらしい。大正解。」さっそく次の質問に移る。「では、日本において妊婦さんが HIV 陽性である可能性はどれくらいだろうか？」質問の意図を見失ったような表情をする学生も現れる。

続けて尋ねる。「100人に1人？」教室のほとんどの学生が首を振る。「1000人に1人？」首をかしげる。「5000人に1人？」うなずく者も現れる。

「よし。では仮に HIV 感染者の頻度を多めに見積もって、1000人に1人だと仮定しよう。この数字を前提に、この2×2表を埋めてみよう。まず、日本の妊婦さんが1000人いたと仮定すると、そのうちに1人 HIV 感染者が存在することになる。感度は99.5%だから、この感染者の検査結果はまず陽性になる。一方、非感染者は999人で特異度は99.5%だから、偽陽性者が4、5人見られる。」こう説明しながら以下のように表を埋める。

	HIV 感染	HIV 非感染
検査陽性	1	5
検査陰性	0	994

「さて、検査陽性である人の HIV 感染の確率は何%になる？」

「6分の1？ 20%以下!？」学生は、この事実が驚く。キツネにつままれたような表情、疑い深いまなざし、教室の雰囲気が変わる。

検査結果を解釈するときの有病率の重要性に気付く機会となる。これに続いて、HIV 感染の危険因子を知りその有無を確認することがどれほど重要か、安易に検査結果だけで診断することはどれくらい危ういか、という議論にまで進むことがある。

自分の受けた教育を思い返しながらか、「こんな具合に教えてもらえたから、検査結果を解釈する重要性に気付くことができた」と思えた一瞬を、学生と共有しようとしている。

⁷ 福岡敏雄:「疑問を作る」総論. EBM ジャーナル 2: 695-700, 2001年

⁸ Urban MA, Reichman RC: Human immunodeficiency virus infection. In Black ER, Bordley DR, Tape TG 編 “Diagnostic Strategies for Common Medical Problems” 2nd ed. Philadelphia USA, American College of Physicians 1999, pp269-279.

EBMを支える情報の専門家のためのワークショップ
2002.12.16-18

システムティック・レビュー概論

津谷 喜一郎

コクラン共同計画とシステムティック・レビュー —EBMにおける位置付け—

津 谷 喜一郎

The Cochrane Collaboration and systematic review —its role in EBM movement—

Kiichiro TSUTANI

はじめに —EBMとコクラン共同計画をめぐる 1990年代の動き—

1990年代は、「エビデンスに基づく医療」(evidence-based medicine: EBM)とコクラン共同計画(The Cochrane Collaboration)が急速な進展をみせた時期である。日本を含めたその動きをFig.1に示す。

EBMという用語は1991年のGuyattの論文¹⁾にはじまるとされる。それが世界的に用いられるようになったのは、1990年代後半で、特に、Sackettらの“Evidence-based medicine: how to practice and teach EBM”が1997年に発行された後である。

一方、コクラン共同計画は、1992年からスタートし、第1回のコクラン・コロキウムは1993年にロンドンで開かれた。その後、コクラン・コロキウムは毎年世界各地で開催されており、またコクラン共同計画の主たる活動であるコクラン・システムティック・レビューもこの回数多くなされるようになった。

日本では、当初、疫学、臨床疫学、臨床薬理学、情報医学などの領域の研究者、その他、特に欧米に留学したものの間でこれらの動きが知られていた。1994年10月1-4日のカナダ・ハミルトンでの第2回コクラン・コロキウムに参加した者を中心として、日本でのネットワーク(Japanese informal Network for the Cochrane Collaboration: JANCO)が同年11月5日に設立された。翌1995年10月にNifty Serveの“FDRUG”というフォーラムに「コクラン共同計画」という会議室が立ち上がった。その後この会議室は1999年2月24日に「コクラン共同計画/EBM」と名前が変えられた。

日本でのEBMの実際の動きはFig.1に示した1997年12月7-8日の名古屋での第1回EBMセミナーが突破口で、その後数多くのセミナーやワークショップが開催されるように

なった。この年6月には厚生省の「医療技術評価の在り方に関する検討会」報告書が発表され、また、11月には「コクラン共同計画資料集ver.2.0」が出版されている。こうした意味で1997年が日本のEBM元年といってもよいだろう。翌1998年3月にはSackettらの本の日本語訳が発行された。

この経緯からわかるように、日本では「コクラン共同計画」という名称がEBMよりも先に伝えられ広がった形になったのである。

先のフォーラムの会議室の名称の追加と変更は、日本でのEBMの認知の高まりにつれて、本来の「コクラン共同計画」の会議室でのEBM全般に関する発言が増えたことに応じてなされたものであるが、ここにみられるように、コクラン共同計画とEBMは、日本では「一固り」になって認識されることが多く、一部に混乱を生じた。

そこで本稿では、まずEBM movementの中でのコクラン共同計画の位置付けを明らかにし、ついでシステムティック・レビューに対するニーズを確認し、コクランのアイデアが具体的にどうコクラン共同計画として結実したかについて述べ、最後にシステムティック・レビューの実際とコクラン共同計画の運営について紹介する。

1. コクラン共同計画とEBMの位置付け

EBMのテキストを開くと、Fig.2のように、step 1: 問題の同定、step 2: 情報の収集、step 3: 情報の吟味、step 4: 患者への適応、という説明がなされる。すなわちEBMは、目の前の患者の問題の同定から始まり、その患者についての問題解決を目指した一連のプロセスである。

一方、エビデンスを中心に全体の流れをみると、Fig.3のように、エビデンスを「つくる」局面、「つたえる」局面、「つかう」局面、の3つに分けられる²⁾。エビデンスを「つくる」(創る, generate)するのは、臨床試験(clinical trial)を中心とした臨床研究(clinical research)である。臨床試験は、(1)評価のために、(2)人を用いて、(3)意図的に開始される、(4)科学的実験、と定義される。

臨床研究は、前向き(prospective)のデザインや後向き

東京医科歯科大学難治疾患研究所・情報医学研究部門(臨床薬理学)

Global		Japan	
The Cochrane Collaboration	EBM	EBM	JANCOG
1991	March/April	Guyatt. <i>ACP Journal Club</i>	
1992			
1993	15-16 Oct.	1st Cochrane Colloquim (CC) at Oxford	
	3 Nov.	<i>JAMA User' guides</i> to the medical literature series started	
1994	1-4 Oct.	5	<i>JAMA</i> 日本語版 医学文献を読むための 「利用者の手引き」掲載開始
		11.5	第1回 JANCOG 会議(東京)
1995	4-8 Oct.	3rd CC at Oslo	11.5 第2回 JANCOG 会議(東京) 12.3 第1回 SRWS(東京) 12.3 「コクラン共同計画資料集 ver.1.0」
1996	19-23 Oct.	4th CC at Adelaide	12.1 第3回 JANCOG 会議(東京) 12.1 第1回 HSWS(東京)
1997		“EBM” by Sackett “EBHC” by Muir Gray	6.27 「医療技術評価のあり方に関する検討会」報告書 11.15 「コクラン共同計画資料集 ver.2.0」 (別府, 津谷, サイエンティスト社)
	20-24 Oct.	5th CC at Amsterdam	12.7 第4回 JANCOG 会議(名古屋) 12.7-8 第1回 EBM セミナー(名古屋)
1998	22-26 Oct.	6th CC at Baltimore	3.10 「EBM」(訳: 久繁, 吉富製薬より配布) 12.19 第5回 JANCOG 会議(自治医大) 12.19-20 第2回 EBM セミナー(自治医大)
1999			1.10 「EBM 実践ワークブック」(名郷, 中山書店) 2.6 第2回 SRWS(東京) 2.26 「効果と効率」(訳: 森, サイエンティスト社) 3.23 「医療技術評価推進検討会」報告書 3.25 「EBM」(訳: 久繁, じほう) 6.1 「EBM 実践ガイド」(福井, 医学書院) 9.18 第3回 SRWS(名古屋)
	5-9 Oct.	7th CC at Rome	9.30 「EBHC」(訳: 久繁, 山之内製薬より配布)
2000			1 <i>EBM</i> ジャーナル発行開始 1.15 第6回 JANCOG 会議(名古屋) 1.15-16 第3回 EBM セミナー(名古屋) 5.1 「EBHC」(訳: 久繁, じほう)
	25-29 Oct.	8th CC at Cape Town	9.10 「EBM のための情報戦略」(中嶋, 中外医学社)

(EBM: evidence-based medicine, EBHC: evidence-based healthcare,
SRWS: systematic review workshop, HSWS: handsearch workshop)

Fig. 1 コクラン共同計画とEBMの世界と日本の動き

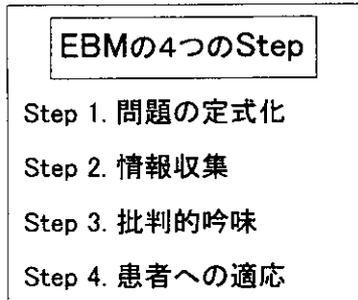


Fig. 2 EBMの4つのstep

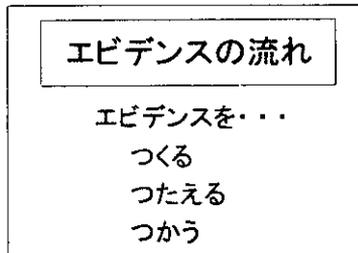


Fig. 3 エビデンスの流れ

(retrospective) のデザインをとる「縦断的デザイン」(longitudinal design) と、時間の推移を含まない「横断的デザイン」(Cross-sectional design) とにわけられる³⁾。これらの臨床研究は広く疫学研究 (epidemiological study) と称されることもある。すなわち、臨床試験は臨床研究のうち、前向きな研究デザインをとるものの一種である。また、臨床試験のうち、日本で医薬品や医療器具などの承認申請のために行われるものが「治験」である。治験に対するびったりした英語はないが強いて訳すと“registration clinical trial”となる。この3者の関係はFig. 4のようになる。

一つ飛んで「つかう」をまず先に説明しよう。エビデンスを使うユーザーは幅広いものである。大きく3つのカテゴリーに分けられる。

第1は、臨床の場の医療従事者である。医師、薬剤師、看護師、保健婦、鍼灸師、その他多くの職種が存在する。彼らは、目の前の個々の患者の問題を解決しようとする人々である。

第2は、集団を対象とする医療従事者である。保健医療や薬事などに携わる行政官、臨床ガイドラインを作成する人々、製薬企業や医療機器の企業で開発などに携わる人々。個々の患者を対象にではなく、マスを対象とし、その健康状態の介入 (intervention) に関わる者すべてということになる。こう考えれば、医療メディアや一般メディアで編集などに携わる者もこのカテゴリーに入る。

第3は患者である。予防や健康増進ということも考えれば一般市民すべてがここに入る。保健医療の各種の介入の受け手であり、税金や各種保険料を払って保健サービスを買う人 (purchaser) でもある。患者はある介入を受けることを1人で意思決定することもあれば、患者グループとして集団で判断することもある。

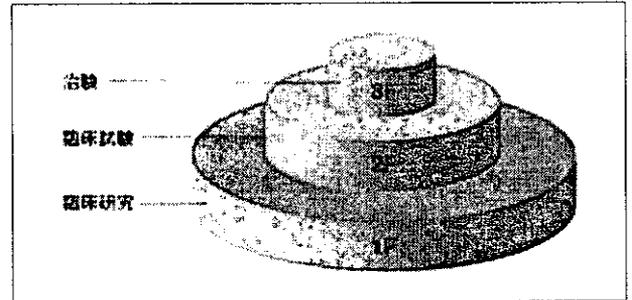


Fig. 4 「治験」と「臨床試験」と「臨床研究」の関係 (中野重行, 治験推進のための環境とその未来, Capsule 2000; No. 66: 18-21より)

さて、エビデンスを「つくる」と「つかう」の間に入るのが、エビデンスを「つたえる」局面である。医師や医療従事者に対し、彼らがどのような情報を利用しているかという調査がいくつかなされている。それらは一般に彼らにとっての「情報源」と称されるが、情報の「源」であるその「出所」と、それが伝えられる「メディア」や「手段」などに分けて提示されることもある^{4,5)}。この「つたえる」局面の「質」の研究はこれまであまりなされておらず研究が遅れた分野であった。そうした状況に対し1990年代に登場したのがシステマティック・レビューを行うコクラン共同計画といえる。

2. システマティック・レビューへのニーズ

上記のEBMの4つのステップと、エビデンスの3つの局面を組み合わせるとFig. 5のようになる。すなわち、EBMはエビデンスの流れからいうと、川下に属し、目の前の患者にエビデンスを「つかう」ことを主とする活動である。

ここで、マスを対象として考えれば、「エビデンスに基づくヘルスケア」(evidence-based health care : EBHC) や「エビデンスに基づく医療政策」(evidence-based health policy : EBPH) ということになる。この2つをあわせてEBHとも呼ぶ、さらにEBHと狭義のEBMを含めて広義のEBMと呼ぶこともある。

EBMにしろEBHにしろ、エビデンスの流れからいえば、エビデンスがどこかで「つくら」れ、誰かが「つたえ」るプ

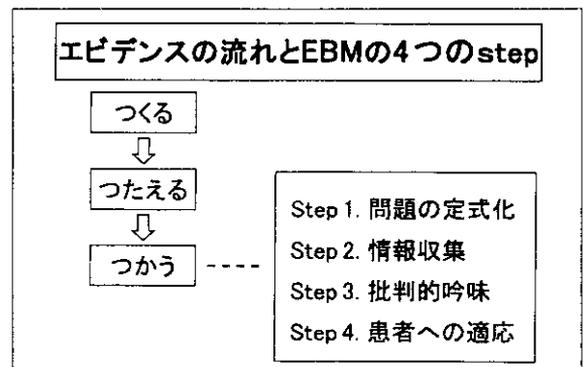


Fig. 5 エビデンスの流れとEBMの4つのstep

ロセスは同じである。違うのは最後のプロセスで意思決定を個人に対して行うのか、集団を対象として行うかということになる。

そして、意思決定に必要なエビデンスは、それが (1) バイアスが小さく、(2) バラツキが小さく、(3) 遅れが小さい、ものがよい。そこでこれを探すことになる。

ところが捜してみると、現代はむしろ情報過多な状況である。ある問題について情報を集めようとする、従来の紙媒体のものに加えて、新しい情報技術 (information technology : IT) の「おかげ」で、情報洪水ということになってしまふことが多い。また量のみならず、そこに書かれている結論がお互いにくいちがっていたり、それぞれのサンプルサイズが小さくて明確な結論となっていなかったり、中には、より最近では否定されたような陳旧化した情報がまぎれこんでいたりする。これらの、あふれるほどの情報からその一片をとりだし、それにもとづいて意思決定することはかなりリスクをとまうことである。

つまり状況として、個人がある問題について、「くまなく」情報を探し、それを批判的吟味 (critical appraisal) することは、限られた時間内には不可能ということになる。

ここで Fig. 2 に示した EBM の 4 つのステップをみてみると、EBM の実践における多くの労力は、step 2 のエビデンスを収集することと、step 3 の情報吟味についやされていることがわかる。

そこで、この 2 つのステップをあらかじめ誰かが行ってくれる、すなわち “pre-appraised review” してくれる。またその結果を遅れなく最近の情報技術を用いて届けてくれるシステムが求められる。

こうしたニーズに答えられるものがコクラン共同計画に代表されるシステマティック・レビューなのである。

すなわち EBM は、情報洪水の時代には “pre-appraised review” と “information technology” を用いて、より “user friendly” な形で提供されるべきもので、またそれが可能な時代になってきているのである。

システマティック・レビューは EBM のインフラ

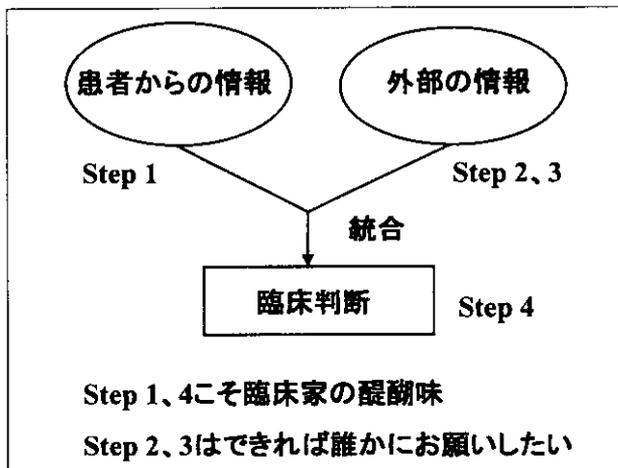


Fig. 6 意思決定者はstep 1とstep 4に専念できるのが理想 (本図は、名郷直樹氏による)

(infrastructure)といえる。これが充分整えば、医療について意思決定するものは、EBMやEBHのstep 1の問題の同定と、step 4の個人や集団へのエビデンスの適応に労力を集中できるようになる。そのあり様はFig. 6であらわされる。

3. コクラン共同計画のアイデアと展開

現在のEBMには3人の父がいるとされる。1人はアーチボルド・コクラン (Archiebald L. Cochrane : 1909-88) である。コクラン共同計画は彼の名前を冠にしたものである。2人目は、アルバン・ファインシュタイン (Alvan R. Feinstein: 1925-) である。EBMの方法論的裏づけとなる臨床疫学 (clinical epidemiology) の開祖の一人ともいえる人である。米国・イエール大学の内科の教授で大変頭のよい人である。頭がよすぎて他人があまりついていけないというような感じの人である。3人目はデビット・サケット (David L. Sackett : 1934-) である。

コクランはすでに亡くなっているが、ファインシュタインとサケットを比べてみると、ファインシュタインは冷たい感じがして「分かる人には分かる」といった風であるのに対し、サケットは、彼の人柄と組織能力により、ある面理解するのが難しい臨床疫学に“EBM”という心地よい「衣」を着せて、世界に広めることに成功したといえよう。

さて、コクラン共同計画が名を冠するコクランは3つのことをいっている。

第1は、「有効な治療はすべて無償で」 (all effective treatment must be free) である。イギリスの国民保健サービス (National Health Service: NHS) は1948年に始まり世界各国の保健サービスの一つのモデルとなったものであった。このシステムは保健サービスが税金でまかなわれるものであるが、そこで一種の政治的スローガンとして「治療はすべて無料で」 (all treatment must be free) というものがあつた。ここでは単純に「治療」と表現されているが、予防やリハビリなどを含みすべての保健サービスとみなしてよい。

このスローガンに対しコクランは反対したのである。彼は「治療はすべて無料」ではなく「有効な治療はすべて無料」 (all effective treatment must be free), つまり「有効」 (effective) なもののみが無料であるべきといったのである。ここには、限られた医療資源をより効率的 (efficient) に用いようという彼の思想があらわれている。効かないものにお金をつかうのは無駄というものだ。この彼の思想は1972年のかれの名著 “Effectiveness and Efficiency” [森 亨訳「効果と効率」, サイエントリスト社, 1999] によくあらわれている。では、「有効」であることを立証するにはどうしたらよいだろうか。

第2は、「ランダム化臨床試験が重要である」である。つまり、ランダム化によって群分けにあたっての選択バイアス (selection bias) を避け、両群の背景因子のバランスをとってこそ、正しい比較ができ、正しい評価ができるとしたのである。これには元々イギリスが世界で最初のランダム化比較試験 (randomized controlled trial: RCT) を行った国であること、さらには、イギリス人の経験論哲学や懐疑主義が反映

している。

第3は、「各トピックごとに、それぞれのランダム化比較試験を、定期的にクリティカルにまとめていないことに関して、われわれ専門家は批判されるべきだ」(It is surely a great criticism of our profession that we have not organised a critical summary by speciality or subspeciality, up-dated periodically, of all relevant RCTs)⁹。

つまり、1)すべてのRCTから、2)よいものだけを、3)まとめて、4)遅れなく、5)必要な人に届けるべきだ、ということである。確かに、どんなによいものがあったとしてもそれが見落とされていたのではなんにもならない。RCTにもよいものもあれば悪いものもある。RCTの結論がくいちがったり、はっきりしない時にはそれらがまとめられていなければ意志決定できない。その結果がおくれて届けられたのでは、意思決定するものは困るというわけだ。

Fig. 7はコクラン共同計画のロゴである。ここに、「Preparing, maintaining and promoting the accessibility of systematic reviews of the effects of healthcare interventions」とある。「システマティック・レビューを、つくり、手入れし、アクセス性を高める」、つまり流通を強化しようというものだ。

このロゴは、新生児の呼吸窮迫症候群 (respiratory distress syndrome : RDS) という病気があり、呼吸不全で亡くなることがあるのだが、分娩中にステロイドを投与すると、それが予防できることが明らかとなったメタ・アナリシスをベースにしている。Fig. 7の中央の縦線はオッズ比1、つまりプラセボと同じしか効かない、左へ行くほど介入、つまりここではステロイドが効くということなのだが、それぞれのトライアルでは結論がはっきり分らなかった。そこで

システマティック・レビューを行ってみたところ、下にダイヤ形で示すように効くことが明らかとなったというものである。その図をThe Cochrane Collaborationの2つの「C」で囲んだデザインが、このロゴである。

このコクラン共同計画を始めた人はIain Chalmersというイギリスの産婦人科医である。Fig. 8に1998年10月にロン



Fig. 7 コクラン共同計画のロゴ



Fig. 8 RCT50周年記念シンポジウム (1998.10.29-30, 後列右から2人目がDr. Chalmers, その左が生物計学者のDr. Peto)

ドンで開催されたRCT50周年シンポジウムでの写真を示す。彼はコクランの弟子で、WHOの協力も得て10年以上こうしたシステマティック・レビューを周産期領域で行っていた⁷⁾。これが大変有益なプロジェクトであることが分かり、すべての領域に広げ1992年からスタートしたのがコクラン共同計画である。

4. システマティック・レビューの実際とコクラン共同計画の運営

システマティック・レビューの具体的な方法は、Fig. 9に示すステップとなっている。第1に、研究テーマ、リサーチ・クエスチョンを正しく設定する。答えられないようなリサーチ・クエスチョンを立て、何とかならないかということも多いのだが、回答可能 (answerable) なリサーチ・クエスチョンを立てることが大事である。第2に、研究をもれなく収集する。出版されていないトライアルを含めて収集する。第3に、各研究の妥当性の評価をする。第4に、それをアブストラクトフォームに要約する。第5に、メタ・アナリシスによる統計学的解析を行う。これは現在ソフトができてから、割合容易にできる。第6にその結果を解釈し、第7、最後に、編集して、定期的に更新して届ける、というわけである。

ここでメタ・アナリシスと、システマティック・レビューという用語は、混同して使われることが多いのだが、コクラン共同計画では、このようなプロセス全体をシステマティック・レビューと呼び、統計学的・数学的なプロセスの部分をメタ・アナリシスと呼んでいる。一般には全体をメタ・アナリシスと呼んでも間違いではない。

1999年にはこうしたシステマティック・レビューやメタ・アナリシスの論文の質を高めるためのチェックリストが作成され公表された⁸⁾。

具体的なコクラン共同計画の運営は、(1)共同レビューグループ、(2)方法論作業グループ、(3)フィールド、(4)コンシューマー・ネットワーク、(5)コクランセンター、の5つのコ

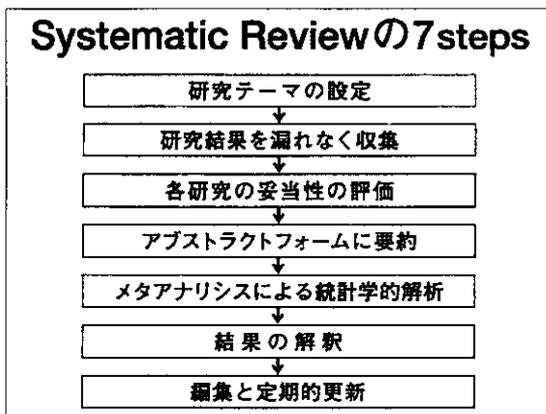


Fig. 9 システマティック・レビューの7つのstep

ンポーネントからなる⁹⁾。これら全体を調製・運営は運営委員会 (Steering Committee) が行っている。全体像をFig.10に示す。各コンポーネントの機能について以下に簡単に述べる。

(1) 共同レビューグループ (Collaborative Review Group : CRG)

2000年現在、基本的には疾患領域別の約60のCRGが存在し、コクラン・システマティック・レビューが、作られ、手入れ (maintain) されている。グループは、研究者、保健医療従事者、コンシューマによって構成される。CRGはコクラン共同計画の目的にどのように貢献するかまず簡潔に述べたプランを用意し、グループ作業の計画、コーディネーション、モニタリングの責任者を明示する。またグループが対象とする分野に関連する可能な限り多くの研究を見出し、特定のレジスタに収集する方法、そしてそのレジスタにある研究を用いてレビューを継続的に作成する責任者も同時に明らかにする。

(2) 方法論作業グループ (Methods Working Group : MWG)

コクラン共同計画が用いる多様な研究成果の統合手法はまだ揺籃期にあり、急速に発展しつつある。MWGは新しい方法論を開発し、コクラン共同計画に対しシステマティック・レビューの妥当性や精度を改善する方法を提言する役割を担う。現時点では「適応性と勧告」「保健経済学」「情報検索」「前向きメタ分析」など11グループが活動中である。

(3) フィールド (Field)

フィールドは疾病別の課題ではなく、ケアの設定 (例えばプライマリ・ヘルス・ケア)、コンシューマのタイプ (例えば小児保健)、介入のタイプ (例えば代替医療) に焦点をあてたものである。フィールドのメンバーは、関連する研究の専門家の情報検索、そのフィールドでの優先度と見通しがレビュー・グループの作業に反映されるような支援、特別なデータベースの編集、外部の関連団体との活動を調整、各領域に関連したシステマティック・レビューへのコメントなどを行なっている。

(4) コンシューマー・ネットワーク (Consumer Network)

コンシューマー・ネットワークは、コクラン共同計画に参加しているコンシューマに情報とネットワークのフォーラムを提供し、世界中のコンシューマ・グループの連絡係となっている。その目的とするところは、コクラン共同計画内部で、コンシューマの発言を促進するような環境を作ることであり、コンシューマ・グループのシステマティック・レビューへの参加、その利用を推進することである。

(5) コクランセンター (Cochrane Center)

1999年3月に中国・成都に中国コクラン・センターが開設され、またいくつかのセンターの統廃合があり、2000年末で13カ国に14カ所のコクラン・センターが活動を行なっている。これらのセンターは上記の各組織の作業を円滑化す