

が立てられた。それまで、「健康と福祉の改善」と多くの人が言ってきたが、科学的な情報無しに医療政策、もしくは臨床行為が行われてきた訳である。つまり、科学的な情報、knowledge-based information（知識に基づく情報）に限界があった。R & D戦略の中では、効果的、効率的な資源を利用するために、正確な評価の情報に基づいて、primary care から医療提供の在り方を変更すべきであると提唱され、健康研究でも実際の応用を目的とした研究を重視し、それに応じて組織のあり方を変えようとしている。基礎研究から問題志向型の研究へ転換しようということ、基礎研究とほとんど同額（簡単な概算で大体 1,000 億円ぐらいの予算）を問題解決型もしくは EBM 型の研究に投入し、情報の伝達のあり方を変えようとしている。

R & D戦略の内容は非常に簡単である。問題を解明して、解決策を幾つか提案する。次に、その中で何が一番望ましいのか、評価・選択する。そして、実際に実行するときの問題点をチェックして、実行に当たってモニタリングをして影響を評価する訳である。また、この R & D戦略の仕組みは諮問委員会と執行部会から構成されている。地域の R & D戦略としては、オックスフォード地域に EBMセンターを設置する、ヨーク地域にはヘルシー・エコノミクスセンターを設置するというように、問題ごとにセンターを各地域に配置している。例えば、EBMについてはオックスフォードが中心となってその普及、適用が図られている。

アメリカの場合、政策的にはほとんど連邦政府の影響力は弱い、AHCPR（保健医療政策研究局）が outcome research（結果の評価）のプロジェクトを 1990 年に設立して以来、25 のプロジェクトを行っている。また、その成果に基づいて実際に 19 種類の臨床指針を作成した。1997 年には Evidence-based Practice Center という EBM 関連の 12 のセンターを作り、急速に普及を図っている。

EBMと医療経済の接点については、今非常に大きな関心を呼んでおり、Lancet 誌や British Medical Journal 誌でいろいろ論争されている段階である。EBMの場合、目の前の患者に最善、最新の医療を提供する。しかしながら、それは個人倫理であって、社会的に限られた資源であるという条件下では逆に非倫理的であると見なされ、様々な問題点が惹起する。Sackett 教授は、臨床の場で実際に個人の医療を考える時もコスト・ユーティリティーを考えていると反論している。しかしながら、全体的に見ると社会的な医療政策レベルでどう考えるかというその連繫について今後十分に検討する必要がある。

医療資源を配分するときにおける未解決な問題として、(1)公正な機会、(2)効用集積、(3)優先順位、(4)民主的プロセス、(5)根拠、が挙げられている。例えば、公正な機会および優先順位をどう保証するのかということが問題とな

ることがある。高齢者と若い人に同じ医療を提供しても、得られる利益は相当異なってくる。利益の大きい方を選択した結果として、若い人のみに限られた医療資源を提供するということが良いのか。また、命の価値を評価するときに、非常に軽微でも広範囲に影響がある場合と、稀れであるが重要な疾患をどのように比較するかという問題がある。虫歯のクラウンと虫垂炎の切除で、オレゴン大学では優先順位は虫歯のクラウンの方が上にランクされた。このように、健康問題においてすら個々人の立場によって資源配分の仕方に相違が生じる。

また、医療資源の配分を考える際、明確な根拠がある場合と根拠がない場合にどう使い分けするのか（例えば、医療資源を必要としているのに、「根拠がないから配分できない」で済ますことができるかどうか）について現在係争中である。これについては未だ明確な結論が出ていないし、このようなことが実際可能かどうかということもまた問題である。

## 6. EBMの挑戦と課題

EBMが提唱されているけれど、本当に有効かがまず問題となる。これについては、十分な検討はまだほとんど行われていないが、実は別の形でEBMが実際に実行されているという評価がなされている。

前述の「医療技術の10～20%しか有効でない」との指摘については、アメリカのテレビ番組でも「80～90%はインチキだ」ということで話題を播いている。しかしながら、現実の医療技術の使われる種類によっては確かにそうであるかもしれないが、実際の患者がEBMに基づいたものを提供されているかが問われている。無効な種類の医療技術が多くても、使われる機会が非常に少なければインパクトはあまりない。また、今までの医療技術の評価は高額値札（ビッグチケット）のテクノロジー、すなわち至極高額の技術にしか焦点が当てられていなかった傾向がある。しかも、評価するために行った文献検索も単純である。Sackett教授はアームチェア（安楽椅子）でやっているから問題だと指摘している。ここでは一般内科での評価例を述べる。

Ellis（Sackett教授の同僚）らが大学病院の一般内科入院治療で評価した結果（1995年）によると、根拠に基づくものが8割であり、実質的な根拠が欠如している治療が約2割であった。根拠については、無作為割付臨床試験に基づくものは5割であり、無作為割付臨床試験はやっていないが説得力のある非実験的根拠による治療が3割であった。

大学附属病院の結果だけでなく、一般の診療所のレベルでの報告もある。Gillら（1996年）が郊外の一般診療所で実際にどう行われているのか検討すると、

根拠に基づくものが大学病院と同様に約8割であり、実質的な根拠がない治療は2割であった。根拠の種類は大学病院とやや異なり、無作為割付臨床試験に基づく根拠は3割であり、説得力のある非実験的根拠による診療は5割であった。

これらの報告に対して、専門家から非難が出されている。「文献に基づいて診療をやっている」とか「自分のところで根拠に基づいて行っている」と回答しているが、それが本当に正しいかどうかはほとんど証明されていないという批判である。中には、単なるリップサービスの見せかけだという批判も出されている。これらについて、Sackett教授のEBMのグループは「想定される批判」であると書いている。

並行して、もっと具体的な実行面でEBMの検討がなされつつある。EBMでは、実際に臨床現場で行う場合臨床指針（クリニカル・ガイドライン）が非常に重要な役割を果たす。これまでに3万とか6万件ぐらいのガイドラインが出されている。ただ、エビデンスに基づく指針はそれほど多いとは限らない。しかし、成熟期に入って、それを適用した場合どういう成果が上がるのかを評価する時期に入っている。

GrimshawとRussellの論文（1993年）では、系統的吟味（システマティック・レビュー）を実施すると、臨床指針の有効性について59の論文が把握される。臨床状態に関するものが24編、予防に関するものが非常に多く27編であった。この中で管理（マネジメント）の過程が変わったというのは、59論文中55編（93%）であった。健康結果の変化について調べた論文は11編しかなかったが、実際に患者の健康結果が変化したのは9編（82%）であり、かなり有望視できると著者らは指摘している。

上の結果は将来EBMの実際の有効性について非常に明るい情報を提供している。しかし、Worrallら（1997年）が再度臨床指針の有効性を同様のシステマティック・レビューで評価した結果では、幾分先程の評価成績と異なっていた。この中で吟味された論文は91編であり、そのうち臨床状態に関するものが35編であり、この中の健康結果に関するものは13編あった。この13編の中で健康結果の改善が見られるとした論文は5編（38%）であった。両者の系統的吟味の内容が違うので、一概に同一視することはできないけれど、この38%だとかなり暗い見通しになってしまう。このように、臨床指針の有効性については今後さらなる検討が必要であるように思われる。

EBMは、これまで述べてきたように非常に重要な役割を期待され、1970年代から1980年代にかけて進められてきた“医療の評価”の最終的なまとめ役として今出てきている。今後、提案されている問題にどれほど役立っている

か、正確な評価を与えていくことが重要である。

にもかかわらず、EBM、EBMといくら叫んでもエビデンスがほとんど欠けているのが現状である。したがって、これをどうするかということが当然問題である。グレーゾーンとって、中間的で真理か偽りかどちらともいえないものについて、どういう判断をして、どういう利用の仕方をするかということが課題となる。今一つは、前述の経済的な問題である。集団的・社会的な倫理として資源配分をどうするかについても検討が求められている。また、臨床政策（クリニカル・ポリシー）としてどうするかということが、今後問われてくると思われる。

さらに、EBMのエビデンス、すなわち「実際EBMを実践し、それが本当に良い結果をもたらすかどうか」について、まだ十分に検討がなされていない。これは今後の課題である。また、EBMと言うと個別の医療の中身だけをとっているのではないかという問題がある。健康改善にとって、特に医療政策（ヘルス・ポリシー）の領域になると、健康改善の決定要因は非常に沢山あるが、別に保健医療が主要な要因であるとは限らない。もっと重要な要因として、環境や教育が影響するであろう。したがって、もっと広い視点からEBMの在り方を位置づけることが非常に重要となる。

## II. EBMと診療・教育

### 1. 1980年代の新しい流れ

1980年代前半のアメリカの病院の一般内科では、同僚の医師の多くがハーバードの公衆衛生大学院のマスターコースで勉強し、修士号の学位を取得後、さらに研究を続けるような風潮があった。この1980年代には「臨床疫学」とか「ディシジョン・アナリシス」と言っていた学問領域が公衆衛生大学院でさかんに講義され、これを修得することによって、データを客観的に評価する立場の臨床医のグループが医療界に出てきたかの印象を与えていた。これの担手はプライマリーケア医（アメリカでは一般内科医が主）であり、彼らが臓器別の専門医が出してきた各種データを患者のマネジメントのために本当に役に立つかどうかという視点で、客観的に評価する視点を持ちつつあった。もう一つの動向は、医療界の外側からデータを客観的に見て、本当に自分たちが理にかなった医療を受けているのかという批判的な目が、医療界の外でも出てきたということである。

ところが、最近急に「EBM」という概念が言われ始めた。EBMは臨床疫学と生物統計学、それにコンピューターリテラシーが臨床医にとっても必要であるという意味に解釈することもできるが、名前が変わった途端に世界中に広まったというのが実情である。

EBM (Evidence-based Medicine) という言葉を一番最初に使ったと自称している人はマクマスター大学の Guyatt 博士である。1991年に、わずか1ページの論文が「ACP Journal Club」というアメリカの内科学会の学会誌に出ている。そこでは、従来の診療の仕方と、Evidence-based Medicine を示している。

症例は、70歳の男性。疲労感を訴えて、ヘモグロビンが9.0グラム、臨床的には鉄欠乏性貧血が疑われた。今までの診療ならば、フェリチンとトランスフェリン飽和率の検査法があるので両方やり、その結果に基づいて結果がピッタリであれば鉄欠乏性貧血と判断するし、結果がピッタリしない時はさらに何かで思いつく検査を行うかあるいは専門医に相談するか、また手近の教科書を見て次の手だてを考える。

### 2. EBMの前教育

Guyatt 博士のいう新しい EBM は、同じ症例に対して、まず MEDLINE でもっとも望ましい検査は何かを知るために、トランスフェリンとフェリチンの敏感度、特異度を検索する。そして、それぞれの敏感度、特異度が書かれている文献を取り寄せ、信頼性を吟味する。その上で、敏感度も特異度もまざっているフェリチンのみをオーダーする。検査をする前の時点での鉄欠乏性貧血の可能性（確率）を数値で表して、このフェリチンの敏感度、特異度を組み合わせ、検査後には、鉄欠乏性貧血がどれぐらいの可能性になるかを数値で表わす。例えば、検査をする前は 75% 位の可能性だったのが、フェリチンの検査結果から敏感度、特異度を組み合わせることによって、例えば 90% の可能性になるとか、そのような数値を用いた理詰めの診断であると書いている。

このたった 1 ページの論文に書かれている内容は、現在言われている EBM と比べるとやや幅が狭いように思われる。今日臨床的に我々が考えている手順は以下のようになる。

眼前の患者に何かわからない問題点が出くわした時、文献を何らかの方法で検索する。この場合、臨床的には二つに分かれる。知りたいと思うテーマを扱った論文があるかないかで結果と結論が大きく異なってくる。「ある」場合には、「文献の批判的吟味 (Critical Appraisal)」を行う。結論が全く一致していない文献ばかりのときにはメタ分析が有用であるし、結論が一定の方向で一致している場合には、その結果を眼前の患者に応用できるかどうかという適用の妥当性について判断することになる。ただ、この「文献検索」の場合に、普通は教科書を読むとかいろいろなセカンダリー・インフォメーションを用いることもあるが、通常は MEDLINE の検索をする。

「文献の批判的吟味」のところでは、1980 年代に提唱されてきた臨床疫学的な物の考え方をを使うし、生物統計学の基本を理解しておかねばならない。

第 4 段階の眼前の患者への応用の妥当性判断では、論文で対象としている患者特性と眼前の患者の特性がどれぐらいマッチしているのか、あるいは適用しない方が良いぐらいマッチしていないのかどうかを判断する。患者の価値観、コンプライアンスの予測も含めて、この判断は一筋縄でできるようなものではない。臨床医としてのいろいろな能力とコミュニケーション能力が必要とされる。

一般的には、このような流れで EBM を行う。しかし、データがない場合には、要素的なデータを統合する方法論であるディシジョン・アナリシスとかメタ分析が必要となる。また、これらを自分自身でやることもあれば、検索した文献の中にこれらの方法論を使ったものが入っていて、それを評価することもある。

臨床上の問題点をどのようにしてピックアップするかは、患者をたくさん診る中で覚えていくより仕方がない。学生や研修医と一緒に勉強してきたMEDLINEの検索方法には様々なテクニックがある。例えば、キーワードを入れる場合に field 設定をどうするか、タイトルだけではなくてパブリケーションタイプとか、パブリケーションイヤーなどいろいろな field 設定の仕方がある。「MeSH (Medical Subject Headings)」をいかにうまく引き出して、それを使うかという「Thesaurus」の使い方も知っておれば一層役に立つ。また、「Explode」という方法を使えば、たくさんの言葉の部分を一度に全部カバーできるし、truncation、「AND」「OR」「NOT」の使い方など、基本的な事柄を理解しておく必要がある。知っておくと便利な検索の基礎知識は以下の通りである。

#### (1) Field

文献の特徴がそれぞれの field に記される。TI(タイトル)、AU(著者住所)、SO(論文の雑誌名、巻、頁、年)、PY(出版年)、LA(言語)、CP(出版された国)、AB(要約)、MeSH(所定のキーワード)、TG(チェックタグ)、PT(論文の形式)。

#### (2) 「in」による field の指定

in を使って field 内に特定の語を有する論文を選択できる。特定の著者、雑誌、出版年、論文形式などを指定できる。「=」で指定できる field は in より限られる。

#### (3) MeSH (medical subject headings)

論文の著者ではなく専門の indexer がつけた所定のキーワード。肺癌であれば lung neoplasms である。これにより幾つかの同義語 (pulmonary cancer, lung carcinoma, lung tumor) に悩まず検索できる。検索には通常 free term (text word) の検索と MeSH の検索が使われる。「thesaurus」という辞書から MeSH 検索する方法と「in MeSH」とおいて MeSH field 内の語を指定する方法がある。ハイフンで語をつないでも MeSH 検索するが subheading の指定の仕方では検索できないことがある。MeSH の難点は検索したいテーマにどのような MeSH 語が使われているかが特に難しい。これが分からず検索に失敗することが多い。thesaurus をみるか、有用な文献が多少分かればその MeSH を参考にする。また、最新の薬剤などは MeSH に収録されない。

#### (4) Subheading

MeSH の内容をさらに限定することができる。診断、治療、副作用、病因、病理学、免疫学などに限定できる。thesauras から MeSH 検索するとき subheading をどうするか質問される。これの選択もときに難しい。副作用の情報が AE (adverse effects) ではなく TU (therapeutic use) に入っていたりする。

#### (5) Explode

MeSH は階層構造になっているので上位の語で下位の語を包括的に検索するのが explode である。Prednisolone の治療効果を検索したいとき、外国ではむしろ Prednisone が使用されている。これらの語の上位語である「glucocorticoides, synthetic」を explode すれば prednisolone, prednisone, dexamethasone, methylprednisolone などがまとめて検索できる。thesauras からの MeSH 検索で explode が指定できる。あるいは explodeglucocorticoides synthetic とおく。

#### (6) 「AND」、「OR」、「NOT」などの検索オペレーター

NOT は不必要なものを除外するのに使用できる。検索法の練習で、検索方法による検索結果の違いの原因をみるのに NOT が使用できる。

#### (7) 「\*」 truncation

単語の先頭部分だけを指定し語尾はどのようなものでも検索する。complication, complicated, complicating などまとめて検索しようとするれば「complicat\*」とする。

インターネット上で検索する場合、ウェブスパーかパブメドというのがあるが、実はそれぞれに微妙な違いがあり、同じキーワードを入れても片方では 2,500 件ぐらい出て、片方では 2,700 件ぐらい出てくるということもある。どれがベストかどうかは別として、少なくともこのように文献検索の基本的なテクニックと実情を知っておく必要がある。

過去に心筋梗塞の既往のある人で、エコー上、心臓の駆出率が 30% の患者がいる例を挙げる。何も症状はないけれど、ACE 阻害薬を使う必要があるかどうかという疑問が出てきた場合、上述した explode heart failure-congestive、次に explode angiotensin converting enzyme inhibitors、それから無作為割付臨床試験 (RCT) などのキーワードを用いる。パブリケーションタイプの field に RCT を入れて、それから asymptomatic という四つの組み合わせで、例えば 12 文献がピックアップされたとする。その中で、もっとも厳格な科学的な方法に則ってやっていると思われる Solved study と Save study をまず読み、みんなで批判的



吟味を行う。

E B Mの第 3 段階に当たる批判的吟味の教育では、「研究デザイン」にはどのようなものがあって、どういう特徴、メリット、デメリットがあるかがもっとも重要となる。最近、「無作為化」ではなく「ランダム化」という言葉を使っ  
てはどうかという意見がある。「バイアス」にはどういう種類があるということも勉強しておく必要があるし、「統計手法」も最低限知っておく必要がある。

端的にいうと、「The New England Journal of Medicine」などで、活字が小さく書かれているメソッドのところを読むのが楽しくなれば、大体その人の論文の読み方はあるレベルに達していると評価される。

表 II - 1 研究デザイン

- 
1. 観察（記述）研究 Observational (Descriptive) Study
    - (1) 症例研究 Case Report
    - (2) 群間比較 Group Comparison
    - (3) 患者対照研究 Case Control Study
    - (4) コホート研究 Cohort Study
    - (5) 時系列研究 Time Sequence Study
  2. 実験研究 Experimental Study
    - (1) 無作為割付対照研究 Randomized Controlled Study
      - ・ プラシーボ Placebo、実薬対照 Active Control
      - ・ 盲検 Blind Test
    - (2) 平行デザイン Parallel Design と Crossover Design
    - (3) 逐次的デザイン Sequential Design
    - (4) 一患者での無作為化試験 N-of-1 Randomized Study
  3. 統合的研究 Synthetic Study
    - (1) メタ分析 Meta-analysis
    - (2) 診療ガイドライン Practice Guideline
    - (3) 臨床決断分析 Clinical Decision Analysis
    - (4) 費用効果分析 Cost-effectiveness Analysis
    - (5) 費用便益分析 Cost-benefit Analysis
- 

「研究デザイン」（表 II - 1）については決まった分類方法はないようであるが、「観察研究」か「実験研究」か、それからデータを統合するようなタイ

プの研究かの三つになる。「実験研究」の中では、無作為割付臨床試験が一番バイアスの入りにくいスタディーデザインだということを教える。

バイアスも色々ある（表Ⅱ－２）。論文を読む上では、「選択バイアス」とか「測定バイアス」、スクリーニングでは、「リード・タイムバイアス」とか「レンクス・タイムバイアス」など、患者の問題に応じて、バイアスの理解をする必要がある。もう一つ、研修医も学生も一度理解すると簡単なのだが、当初なかなか理解しにくいのが「交絡因子」である。例えば、お酒を飲む人の集団とお酒を飲まない人の集団で比べると、飲酒者のグループでは肺癌の発生率が高い。でも、飲酒をしている人たちの中で喫煙している可能性も高いことがわかっている。本当は喫煙のほうが肺癌と関わっているために、一見飲酒が肺癌発生率と関わっているように見える。飲酒と喫煙の関連は正しくて、喫煙と肺癌発生率の関連は正しいのだけれど、表面に出てきた飲酒と肺癌の関係は正しくない、ということになる。このような因子を交絡因子という。

表Ⅱ－２ 文献の評価と解釈に役立つバイアス

---

|            |                     |
|------------|---------------------|
| 1. 選択バイアス  | 5. 出版バイアス           |
| 2. 情報バイアス  | 6. 入院率(Berkson)バイアス |
| 3. 観察者バイアス | 7. リード・タイムバイアス      |
| 4. 面接者バイアス | 8. レンクス・タイムバイアス     |

---

幾つかのテーマ別チェック項目が出版されている。それらをあまり鵜呑みにしてチェックするのは余り実際的でないように思われる。例えば、患者の治療法を扱った文献でランダムに割り当てられているかとか、どうしてこういうのをチェックしなければだめなのかということ自分の頭で考えながら、項目が思い浮かぶようにしないと身に付かない。

### 3. EBMの診療の実際

EBMの実践活動は、眼前の患者にEBMで得た証拠を応用する場合に、どういう点を考えるべきかである。文献で対象となった患者群と自分が見ている患者とがどれぐらいフィットするのか、またはデータを応用してはならないぐらいの大きな隔たりのある患者群なのかという判断を行う必要がある。

風邪の患者に抗生剤を使った方がいいかどうかについて日本中の医者が、風

邪の患者が来たときに1回1回やるのは非常に不合理な話である。このような頻度の高い問題については、皆が使えるような診療ガイドラインを作ってしまうのが当然考えられるやり方である。歴史的には、1970年代のカナダの癌スクリーニングのガイドライン作成時に、「診療ガイドライン」を科学的な評価に耐え得るものを作成するためにはどうしたら良いかについての基本的な方法が確立された。その後、1980年代に入って、アメリカのUS Preventive Service Task Forceで、スクリーニングの有効性を評価した。例えば、肺癌検診のためのX線撮影はまったく無効とか、169項目について評価している。臨床上、遭遇する非常に頻度の高い問題については、このような診療ガイドラインをつくる作業が既に行われており、それらの有効性が最近評価されるようになっている。

#### 4. EBMの教育の実際

日本におけるEBMの教育として、医学教育学会で臨床疫学のカリキュラムが作成されている。その中に、「臨床疫学の基本的概念を理解する」とか「EBMと臨床疫学の関係を理解する」とか、「患者の有する問題を解決するプロセスで、臨床疫学的概念を活用する能力を身につける」という「一般目標」と「行動目標」を作って、授業90分ごとにどのような内容を扱い実習をすれば良いかが記されている。

「総論」では、バイアスや内的妥当性・外的妥当性、文献の批判的吟味、診断検査の検査前確率という概念や“gold standard”、リスクと予後、治療、無作為割付比較試験、目隠し評価、NNTという統計指標の表し方、メディカル・インジャーリー、予防、それからチャンスイベント等統計学の基本、メタ分析とかQOLの評価の仕方などが挙げられている。また、費用効果分析も含めてディシジョン・アナリシス、コンピューターを用いたMEDLINE検索の講義・実習の他に、実際に文献を渡してそれを批判的に吟味するという90分授業で15コマでカバーできるようなものが提示されている。

ハーバード大学では、1986年から始まった「Clinical Effectiveness Training Program」というプログラム（表Ⅱ-3）があり、夏休みに集中講義として「臨床疫学」、「Biostatistics」、「Health Services Research」などを教えている。また、通年の授業が二つだけあり、第一学年では、Social Medicine and Clinical Epidemiologyが通年で毎週1回2時間ずつ行われおり、EBMそのものを1年目からかなり集中的に教えている。こういう動きからも理解できるように、EBMが正規のカリキュラムにどんどん入りつつあるというのが実情である。

表Ⅱ－3 ハーバード大学 Clinical Effectiveness Training Program

---

1. Study design
  2. Multivariate statistical techniques
  3. Measurement of comorbid medical conditions
  4. Assessment of severity of disease
  5. Measurement of health status and quality of life
  6. Definition of "appropriateness" of procedures
  7. Measurement of costs
  8. Determination of cost effectiveness
  9. Measurement and improvement of quality of care
-

### Ⅲ. 日本における EBM の現状

日本では、一部の大学で EBM を実施する試みがなされているが、学問・教育環境として日本独特のエビデンスを得る過程そのものに問題点があるようである。簡単に、日本のエビデンスの状況と諸外国からのエビデンスを日本に導入することの妥当性について概説する。

#### 1. 日本における科学的“根拠”

日本における“エビデンス”を得る努力はされているのかに関してはまさに悲観的状況にある。一次予防に関する研究では、ごく一部「禁煙」問題などで無作為割付臨床試験が実施されているが、長期的なフォローアップがされていないので必ずしも十分とは言えない。二次予防では、無作為割付臨床試験による“根拠”が全くない。臨床検査については、検査によって正確に診断できるかどうかは中間的な有効性であり、最終的に検査した場合と検査しなかった場合で治療方法、管理方法が変わり、さらに健康改善が認められるか否かを評価しなければならない。その意味で、本当は無作為割付臨床試験が必要であるが、ほとんどないのが現状である。

治療に関して、内科的な治療手技や各種医療技術を組み合わせて、どのような治療法がいいのかという検討はほとんどなされていない。日本で実施された唯一の無作為割付臨床試験を用いた研究として、NIDDM 型糖尿病患者への強化インスリン療法の有効性の検討がある。この結果、強化インスリン療法は非常に優れていることが示されている。これは国際的に通用する論文ではあるが、糖尿病のタイプが日本と外国で多少違うので、それが国際的に有効かどうかは更に検討が必要だと言われている。外科の治療手技についてはほとんどない。特に、低侵襲性の医療（外科的）技術に関しては、外国ではきめ細かい無作為割付臨床試験が行われて評価されているが、日本では皆無である。薬剤に関しては、法律に規則があるためでもあるが、無作為割付臨床試験が行われている。ただ、脳循環の代謝改善薬のように問題があり、質が低いと指摘されている。

無作為割付臨床試験によるものをエビデンスと定義すると、世界に 50 万から多く見積もって 100 万あるとされている。日本では、大体 1 万である（もっとも、日本の医学関係のデータベースにはランダムイズド・コントロールド・トライアルとか無作為化比較試験という thesaurus がないので、非常に調べにく

い)。したがって、世界中のエビデンスの 1 %は日本にあるということになるが、これはほとんど医薬品に限られている。日本の薬の消費量が世界第2位と言われているのに比べると、この数値は小さすぎる。

## 2. 外国産のエビデンスに依存的で良いか

日本でのエビデンスが少ない状況について述べたが、では外国のエビデンスに頼って良いかどうか問題となる。これに対しては、現在厚生省で設置している医療技術評価推進会議が検討を進めている。この会議の中で、テクノロジー・アセスメントで評価した根拠をいかに臨床で使うかという枠組みについて重点的に検討しようとしている。

現在のところ、例えば日本でガイドラインをつくらうとしても、利用できる情報がほとんどない。仮にあっても、症例報告が多く、コホート研究などは一部にあるかどうかである。このような時、外国人のデータが人種差を超えて使えるかどうか問題になる。

薬の臨床試験で例にとると、もっとも問題になるのは薬が人種差を超えて有効であるかということである。現在のところ、大体8割から9割ぐらいの薬は人種差を超えて有効であろうと言われており、残りの1~2割は疑問視されているようである。したがって、この辺の検討をどういう枠組みで考えるかが問題となる。もう一つは、もし無作為割付臨床試験ができない場合、学会レベルで大きなコホートを組み少なくとも後向き（レトロスペクティブ）にコホートを設定して、各種交絡要因を調整すれば、ある程度、次善か次々善ぐらいの情報が得られるし、それを適用できるのではないかと考えられる。

発症率とか様々な予後データは、外国とほとんど違う可能性があるので、質が低くとも日本で情報を集めてサポートすることが必要になる。上で述べた薬の例のように、本当に8割ぐらい適用できるのかどうかは個別の問題で検討していかななくてはならない。このため、各領域で重要な課題については（システマティック・レビューやメタ分析を含め）追試のフォローや正確な検討が必要となってくる。日本において、個々の研究者が十分な数の対象者を集めることは難しいので、各学会の中堅クラスがEBMの方法を身につけ、実際に適用できる体制を構築することが急務であるように思われる。

## IV. EBMと予防医学

1995年の終わりに、合衆国予防サービスタスクフォース（The U.S. Preventive Services Task Force）は「臨床予防サービスガイド（The Guide to Clinical Preventive Services）」の第2版を発表した。これは1989年に出版された第1版（U.S. Preventive Task Force, 1989）に含まれていた勧告を大幅に更新させたものである。これらの報告書に含まれる勧告は、プライマリケアに従事する臨床医に対する、臨床環境における効果的な予防的ケアの手引きの重要な情報源となっている。タスクフォースの勧告は多くの保健計画や保健部門に採用され、「国家品質保証委員会（The National Commission of Quality Assurance）」が定めたもののような質の指標に盛り込まれている。

このタスクフォースは、保健社会福祉省の疾病予防健康増進局による一つの試みとして始まった。その使命は、特定のスクリーニング検査、予防接種、カウンセリング療法について有効性（effectiveness）の根拠を注意深くかつ客観的に吟味することだった。さらには、特定の年齢層の男女に対してどのような予防ケアを日常提供すべきかについての勧告を定式化する作業も行った。この試みの重要な目標は、どのような対策が真に効果的（effective）であり価値があるかについての合意を得ることにあった。予防的介入の価値について、臨床医たちは混乱し、保険会社は懐疑的だった。

このタスクフォースは、連邦政府から研究費の支援を受けている非政府関係者--しかし政府とは独立ではない--専門家パネルによって構成された。2つめのタスクフォースは、家庭医学、内科医学、産婦人科学、小児科学、予防医学を代表するメンバー10人とスタッフ4人（医師13人、PhD 1人）で構成され、全員が批判的研究評価の専門家だった。このグループは新たなトピックを検討し、重要な新しい研究が存在する重要なトピックを再検討し、当初の手引きを全面的に改訂した。このタスクフォースは、カナダにおける同等のグループ、「カナダ定期健康診断タスクフォース（The Canadian Task Force on the Periodic Health Examination）」と密接な連携を保って活動し、一定の小委員会では文献や背景報告に関する検討作業を共有した。

### 1. 方法論

このタスクフォースの方法論の特徴は、臨床における個々の予防サービスの

有効性 (effectiveness) を決定するにあたって、根拠に基づく明示的な方法を採用した点にある (Woolf et al, 1996)。このアプローチの重要な要素として、有効性を推論するにあたって、関心のある評価項目 (endpoint) とそれに必要な条件を明示的に定義すること (「因果パス (causal pathway)」を用いる、図1) が含まれる。さらには、関連研究を系統的検索、個々の研究の質を評価するための基準、入手できる根拠の強さと直接関連づけられた勧告を作成するプロセスのための方法も含まれる。

このタスクフォースは勧告を作成するにあたって、専門家の意見や現在主流となっている医療に基づくのではなく、有効性に関する科学的根拠だけを厳格に基礎にする。このアプローチの仕方によって、合衆国の他の公衆衛生サービス機関や専門家の学会とは異なった立場に立つ場面も生じたが、そのことは勧告が、政府機関や学会の特定の行動目標には依存しないものであるという事実を浮き立たせることになった。

このタスクフォースは、それぞれの介入が、中間的アウトカム (たとえばヘモグロビン濃度の改善) ではなく、臨床的アウトカム (たとえば罹患率や死亡率の低下) を改善させるという根拠を必要とした。この基準は長期間の臨床試験や、比較対照観察研究を重視することになるので、一部の専門家委員会で用いられる基準よりも保守的な性格をもっていた。今回採用された考え方は、無症候者に日常用いることを促進する前に、何らかの介入の便益が明確にその害を上回っていることを証明すべきだ、というものだった。

タスクフォースでは、スクリーニングやカウンセリングの便益を確立するのに必要な基準を定めるために因果パスの枠組みを用いた。さらに、ルーチンのスクリーニングが「下流に」もたらす影響を、スクリーニング検査それ自体 (実行しやすさ、リスク、偽陽性、偽陰性結果) と、臨床症状が出る前の早期の疾患患者の治療増加の両面から明示的に検討した。カウンセリングによる介入については、(運動と心疾患のような) 健康アウトカムに独立に結びついた特定の行動と、ルーチンのカウンセリングが特定の行動変容をもたらすという、もっと弱いその他の根拠を検討した。

## 2. 根拠のレビューと評価

タスクフォースは Medline 検索を行って、因果パスの個々のステップを扱った発表済みの文献を探した。関連のある論文を特定し、その研究の質の評価を行い、研究デザインの種類に従って分類した (表IV-1)。ここで示した階層構造は、無作為割付比較対照試験と比較して観察研究にはより多くのバイアス



が含まれることを反映しており、この要素だけで根拠の強さの重みを定めた。タスクフォースではさらに、研究の質と数、結果の一貫性、知見をプライマリケアの環境に一般化可能かどうか、といった点も考慮した。こうして、支持する根拠の全体的な質を反映させた、日常的にそのサービスを実施すべきだという「強い」根拠のあるクラスAの勧告から、サービスを止めるべきという「強い」根拠のあるクラスEの勧告まで、段階付きの勧告が作成された。

表Ⅳ－1 根拠の程度と勧告の強さ

| 質      | 根拠の水準  |
|--------|--|
| I      | 少なくとも1つの適切な無作為割付比較対照試験による根拠がある   |
| II-1   | よくデザインされた比較対照試験による根拠がある  |
| II-2   | よくデザインされ、できれば複数のセンターや複数の研究グループによるコホート研究もしくは症例対照研究によって得られた根拠がある   |
| II-3   | 介入の有無の違いをみた複数の時系列データから得られた根拠がある。対照が設定されていない研究における劇的な結果（例えば1940年代に導入されたペニシリン治療の効果のようなもの）もこの段階の根拠と見なしてよい |
| III    | 臨床経験に基づき、尊敬されている権威の意見  |
| 勧告の強さ* |  |
| A.     | 当該条件を定期的健康診断において特異的に考慮すべきという勧告を支持する十分な根拠がある  |
| B.     | 当該条件を定期的健康診断において特異的に考慮すべきという勧告を支持するかなりの根拠がある   |
| C.     | 当該条件を定期的健康診断に含めたり除外したりする勧告を出すには根拠が不十分であるが、別の理由から勧告が行われるかもしれない  |
| D.     | 当該条件を定期的健康診断から除外すべきという勧告を支持するかなりの根拠がある   |
| E.     | 当該条件を定期的健康診断から除外すべきという勧告を支持する十分な根拠がある  |

\*カナダ定期健康診断タスクフォースから引用して改変

クラスAの勧告のほとんどは、複数のランダム化試験によって明瞭な便益が示されていることを反映しているが、たとえば子宮頸癌のスクリーニングに対する勧告のように、さまざまに幅広い観察研究に基づいているものもある（表IV-2、表IV-3）。タスクフォースによって評価された予防サービスの多くがクラスCの評価を受けたのは、そのサービスをルーチンとして実施するのを推奨したり、否定したりするのに十分な根拠が入手できなかったことを反映している。クラスCの勧告は、さまざまに異なった状況から生じるが、最も普通には、(1)入手できる根拠に限界があったり、(2)一貫性がなかったり、(3)質が悪かったりすることによる。

表IV-2 平均的なリスクをもつ女性に対する勧告（米国）

---

#### スクリーニング検査

- ・身長と体重
- ・血圧（2年に1度）
- ・総コレステロール（45歳以上）
- ・パプ塗抹標本試験（少なくとも3年に1回）
- ・1、2年に1度のマンモグラム（50歳以上）
- ・1年に1回の便潜血反応試験、もしくは定期的なS字結腸鏡検査のいずれか、または両方（50歳以上）
- ・問題飲酒の評価（標準化されたスクリーニング問診票の結果を用いる）
- ・風疹の血清学的検査、もしくはワクチン接種歴：妊娠可能な年齢の女性

#### 予防接種

- ・破傷風-ジフテリアの追加接種（10年に1回）
  - ・風疹：妊娠可能な年齢で感受性のある女性
  - ・インフルエンザ（毎年）：65歳以上、および慢性疾患にかかっている青年
  - ・肺炎双球菌ワクチン（1回）：65歳以上、および易感染性宿主の個人
  - ・B型肝炎：以前にワクチン接種を受けていない若い女性
- 

タスクフォースは、根拠が不十分な状況では、臨床対応策の指針として他の要因も考慮するのが有用かもしれないと認識した。こうした要因としては、疾病負荷、潜在的なリスクの大きさ、介入に要する費用といったものがあつた。

たとえば、家庭内暴力の日常的スクリーニングが女性にとってよりよいアウトカムをもたらすかどうかを判定するに足る根拠は存在しない(クラスCの勧告)が、家庭内暴力の存在割合が高いこととこの情報の潜在的価値を考えれば、臨床医が通常の成人患者にルーチンに尋ねる既往歴として、虐待に関する直接的な質問を少数追加するよう勧告できるとタスクフォースは考えた。

タスクフォースは、一定のスクリーニング検査については、ルーチンに行うことに反対する勧告も出した。例えば、スクリーニング検査が有効でない根拠がある場合や、重要な便益が得られる根拠がないのにルーチン検査による潜在的な害が明らかに存在する場合である。この後者の問題は、とりわけ、適切な対照比較試験で評価されていない癌のスクリーニング検査に関連していた。癌のスクリーニングで陽性の結果が出ると、多大な不安を起こすとともに、次々に侵襲的な診断手続きや協力的な治療が開始されることがある。タスクフォースでは6種類の癌については日常的なスクリーニング検査の実施に反対している。

タスクフォースの勧告は、コストとコスト対有効性 (effectiveness) の問題を明示的には考慮しなかった。これは、コスト対有効性の分析における基本的な方法論の原理に議論があるためでもあり、また、コストに関する議論をすることで、有効性に関するより基本的な問題から関心がそらされてしまうからであった。「保健と医療におけるコスト-有効性パネル」の最近の報告では、予防的措置のコスト-有効性を評価するための枠組みが提示されている。将来の評価ではコストの問題をもっと直接的に扱うことになるかもしれない。

特定の予防サービスに関する勧告に加え、タスクフォースは、プライマリケア環境における重点とサービス提供についていくつかの一般的結論に達した。

- (1) 患者の個人的健康関連習慣を対象にした介入はきわめて重要である。
- (2) 臨床医は、検査のオーダーや予防サービスの提供にあたって選択的でなくてはならない。
- (3) 臨床医と患者とが共同して意思決定をしなくてはならない。
- (4) 臨床医は、とりわけケアへのアクセスが限られている人に対して、あらゆる手を尽くして予防サービスの提供を試みなくてはならない。
- (5) 一部の健康問題については、臨床的な予防サービスよりも、地域レベルの介入のほうがより有効性が高いかもしれない。

### 3. 特定の勧告の要約

カウンセリングや化学的予防療法に関する特定の勧告の例を、表2 (女性)

と表Ⅳ－3（女性）に示した。

表Ⅳ－3 カウンセリングと化学的予防療法（米国）

---

薬物使用

禁煙の助言

アルコール/ドラッグの使用

食事と運動

脂肪を制限し、カロリーのバランスを保ち、果物/穀物/野菜を重視する

適切なカルシウムの摂取

定期的な身体運動

外傷の予防

シートベルト（腰と肩）

自動二輪車や自転車のヘルメット

煙検知器

銃などの安全な保管と撤去

STDの予防（ハイリスク行動の回避、コンドーム/殺精剤付きの女性用ペッサリー）

適切な避妊：妊娠可能な女性

歯科保健

歯科医院の定期的受診

フッ素入り歯磨きを用いたブラッシングとフロス（毎日）

化学的予防療法

葉酸を含むマルチビタミン/マルチミネラル：妊娠可能な女性

ホルモン予防投与療法のリスクと便益とを説明：閉経後の女性

---

#### 4. 公衆衛生と予防医学に関する実践指針

上述したように、合衆国タスクフォースはプライマリケア医学に焦点を当てた。最近では、科学的基盤に立脚した公衆衛生実践への取り組みを提供しようという気運が盛り上がってきた。こうした取り組みの一例が、ケロッグ基金が資金援助している、メリーランド州ボルティモアにある「公衆衛生実践に対する指針作成プロジェクト（The Guideline Development Project for Public Health Practice（1995）」の報告に記述されている「結核治療パネル（The Tuberculosis