

平成 12 年度文部省 科学研究費補助金

基盤研究 C(企画調査)

課題番号：12897028

「効率的で良質な医療を目指した病態検査の系統的再評価
の基礎的検討」

報告書

班員

石田 博(川崎医科大学 検査診断学)

北村 聖(東京大学医学部臨床検査医学)

三宅一徳(順天堂大学医学部臨床病理学)

西堀眞弘(東京医科歯科大学医学部附属病院検査部)

松野容子(山口大学医学部附属病院検査部)

目次

. はじめに	2
. 各論	4
-1. 検査医学における Evidence の現状について	4
-2. 検査領域における系統的再評価の実践における問題点について	8
-2-A 文献資料検索 - MEDLINE を対象として	8
-2-B 論文の質の評価について	15
-2-C. 統計的手法 (Meta-Analysis)について	19
-2-D. 出版バイアスについて	24
-2-E. Evidence の普遍性(Generalizability)とガイドライン	25
-3 . EBD データベース	27
【1】調査・検討経過	27
【2】EBD データベースの必要性と目的	27
【3】EBD データベースの構造と構築手法	28
-4. 自施設での Fact について	33
-5. EBD 啓蒙と協力体制について	34
-6 . EBD をとりまく国内外の動向	35
【1】調査経過	35
【2】国内の動向（本研究班の活動を除く）	35
【3】国外の動向	36
【4】EBD 研究における本研究班の位置づけ	37
-7. 今後の課題	37
1. Evidence の作成とその支援	38
2. Evidence の集積	38
3. Evidence の有効利用	39
. 総括	40
. 謝辞	42
V. 資料	43
資料 1	43
資料 2	44
資料 3	45
資料 4	48
資料 5 Summary ROC 曲線についての主な論文	51
資料 6	53
. 参考文献	64

はじめに

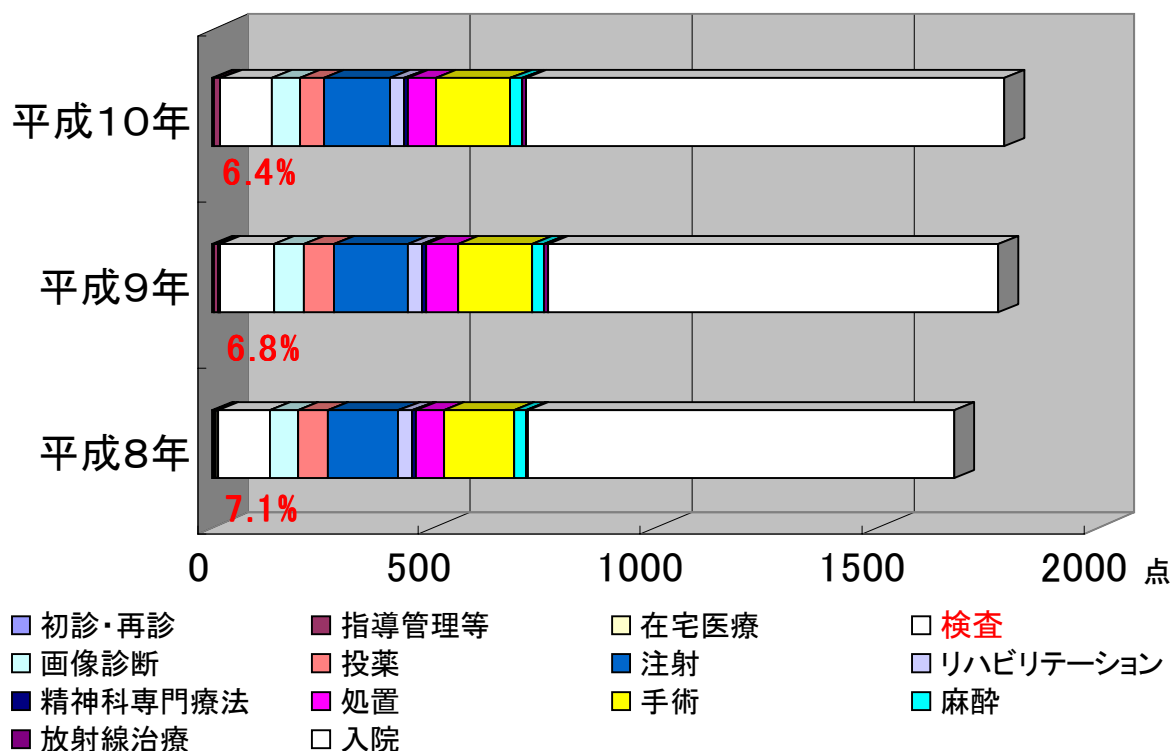
一般臨床あるいは健診において、検体検査は日常的に行われ、疾患のスクリーニングや診断確定、重症度判断、予後判断などに活用されている。また、それらから集積される情報は、予防医学といった公衆衛生的なアプローチや日々の患者マネジメントにおける臨床医学的アプローチの両面で、今日の医療に必要不可欠なものとなっている。一方、検査そのものの診断特性の不完全さから、不必要に検査が実施されたり、重複した検査で同じ情報量しか得られないといった無駄な検査もなされているのが現状である¹⁻²⁾。また、さまざまな新規検査項目の導入と日常検査の高い実施頻度から、医療保険における検体検査の占める割合も約7%を占め(図1)、検査点数の包括化による抑制にもかかわらず、依然として医療資源消費の大きな因子となっている。

そのため、最近では、検査点数の包括化によってその増加が抑制されている。(図1)

図1 入院患者一日あたりの診療点数

(平成11年 社会医療診療行為別調査の概況：厚生省大臣官房統計情報部資料より)

http://www1.mhlw.go.jp/toukei/h11-siryu_8/



このような状況において、良質な検査を選択し、診断効率を高めて DRG/PPS やクリニカルパスの導入にも対応でき、最終的には医療費の削減に貢献できる検査の使い方が模索されてきている。2000 年春には日本臨床検査医学会(旧称：日本臨床病理学会)から「DRG/PPS 対応 臨床検査ガイドライン(二次案)」が発刊された³⁾。また、他の専門学会から提唱されている臨床ガイドラインの中にも、高脂血症のスクリーニングなど検査に関連したものが数多く見られている。

(http://www.guideline.gov/FRAMESETS/static_fs.asp?view=resources.guidel、
<http://www.mi-net.org/>)

これらのガイドラインでは良質な検査項目を求める必要があるが、現状では、その多くが病態生理学的な特徴や専門家のコンセンサスを得るに留まり、それぞれがどの程度の診断特性(感度・特異度)を持つのか、あるいは、その検査が陽性である場合にどの程度、疾患のリスク比が高まるのかといった具体的で客観的な事実(Evidence あるいは Fact*)に基づくまでには至っていない。そこで我々は、診断検査における Evidence の現状と、今後、Evidence に基づいた臨床検査医学：Evidence-Based Diagnosis (EBD)または Evidence-Based Laboratory Medicine (EBLM) を実践していくための問題点について調査、検討を行ったので報告する。

本報告の内容は以下の通りである。

各論

1. 検査医学における Evidence の現状について
2. 検査領域における系統的再評価の実践における問題点について
 - A. 文献資料検索 - MEDLINE を対象として
 - B. 研究論文の質の評価
 - C. 統計的手法 - 診断検査の Meta-analysis
 - D. 出版バイアス
 - E. Evidence の普遍性とガイドライン
3. EBD データベースの確立
4. 自施設での Fact について
5. EBD の啓蒙と協力体制
6. EBD をとりまく国内外の動向
7. 今後の課題
 - ・ 総括
 - ・ 資料

* 本集では、施設が異なっても適応できるような、すなわち外挿性の高い根拠について Evidence という用語を用い、ある特定の施設内において日常的な検査から得られた事実について Fact という用語を用いている。

.各論

-1. 検査医学における Evidence の現状について

現状での検査領域(検体検査・生理機能検査)における Evidence がどの程度あるのかを 2001 年 3 月段階で検討した。検索は、電子的に行える標準的な医学文献データベースである MEDLINE を PubMed により、さらに 2 次峻別資料である The Cochrane Library 2001 issue 1 と Best Evidence 5 を対象として行った。

検索は、Medline と Cochrane Library が index 用語である MeSH (Medical Subject Headings)を用いて以下のように行った。また、Best Evidence では Medline における subheading にあたる分類項目(therapeutics, diagnosis, prognosis など)で検索が可能であることから、その分類項目(diagnosis)での抽出を行った。

MEDLINE で行った検索を下記に示す。検索期間は 1980 年 - 2001 年 3 月である。

(1)診断特性について

Search	Most Recent Queries	Result
#7	Search #4 OR #9	232
#6	Search #3 AND #7	69
#5	Search systematic review	5361
#4	Search #1 AND #2Limits: Meta-Analysis	163
#3	Search #1 AND #2	45104
#2	Search "sensitivity and specificity"[MESH]	101054
#1	Search "Diagnostic Techniques and Procedures"[MESH]	1880620

(2)検査による危険・予後予測について

Search	Most Recent Queries	Result
#9	Search #6 OR #8	684
#8	Search #1 AND #7	529
#7	Search systematic reviewLimits: Review	5361
#6	Search #4 AND #1Limits: Meta-Analysis	155
#5	Search #4 AND #1	49156
#4	Search #3 OR #2	252175
#3	Search Risk[MESH]	247451
#2	Search "Odds ratio"[MESH]	12876
#1	Search "Diagnostic Techniques and Procedures"[MESH]	1880620

検索後、年度別の検体検査、生理機能検査に関する論文をタイトルとアブストラクトから限定した結果は以下(図 2、3)の通りである。

図 2：診断特性についての論文数：Medline からの検索

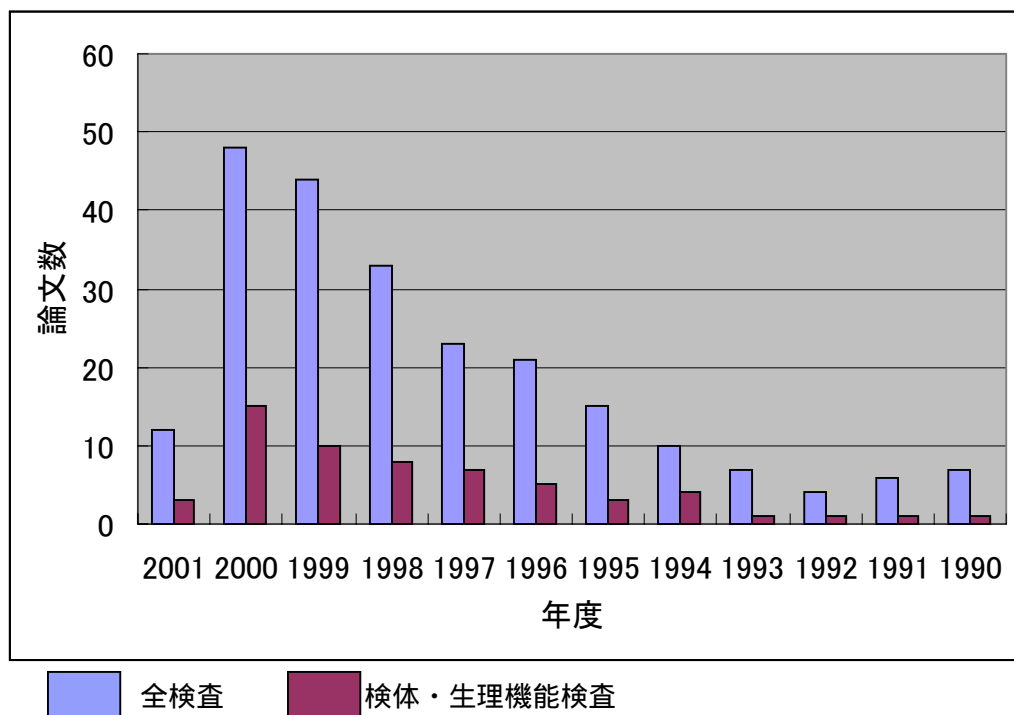
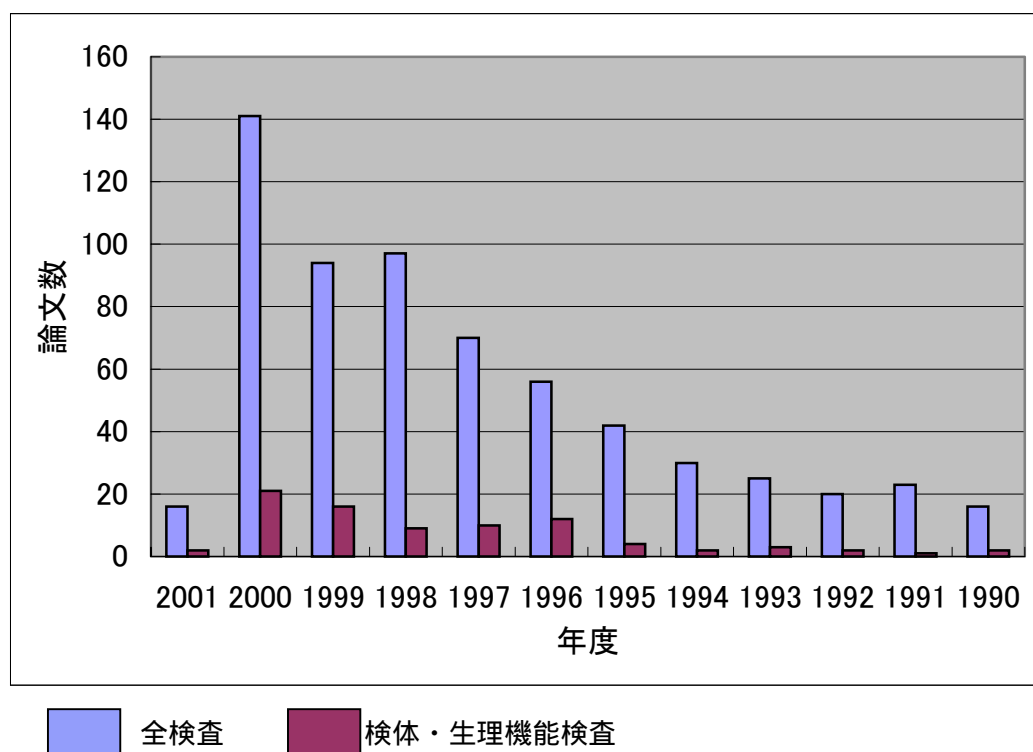


図 3：検査による危険・予後予測についての論文数：Medline からの検索



(2001 年は 3 月まで)

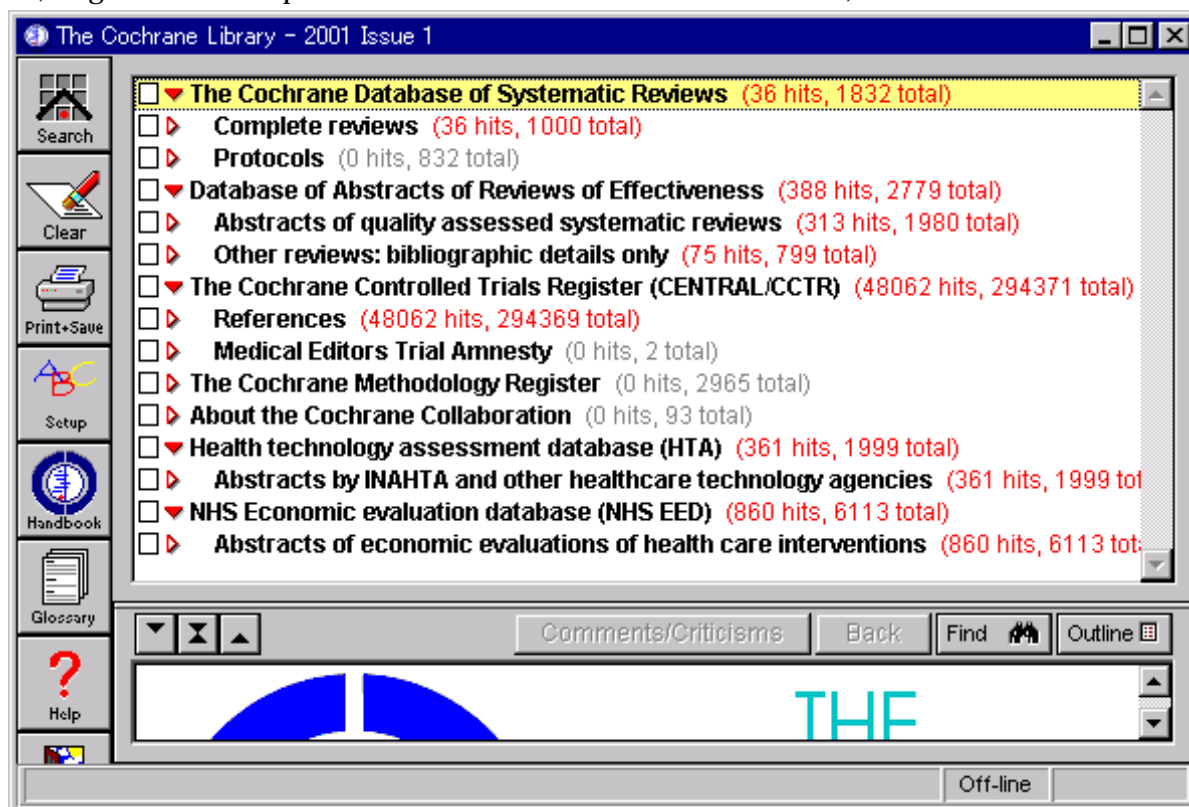
今回は、系統的再評価(Systematic Review : SR)がなされた論文を対象としたが、RCT

を基にした治療関連の SR が一般的になりつつあるのに対して、診断検査に関連した SR は増加傾向にあるものの、現状では少数に留まっている。

また、The Cochrane Library での検体・生理機能検査についての Review についても同様である。まず、下図は、Search メニューで Medline の MeSH 用語である”Diagnostic-Techniques-and-Procedures”によって検索したものである。The Cochrane Library 2001 年 issue1 版で、Cochrane グループによる SR(Complete reviews)総数 1000 件のうち、36 件がヒットし、その中から臨床検査・生理機能検査に絞るとわずかに 5 件であった。また、Cochrane グループとは別になされた SR(Database of abstracts of Reviews of Effectiveness : DARE[†])においては、掲載総数 1980 件のうち、313 件ヒットしたが、その中で検体・生理機能検査関連は 46 件であった。さらに、”Sensitivity and Specificity”での用語に検索した結果は、Complete reviews には hit がなく、SR(DARE)では 112 件ヒットし、該当するものは 46 件であった。

図 4 : Cochrane Library による検索結果一覧

(Diagnostic-Techniques-and -Procedures について検索した内容)

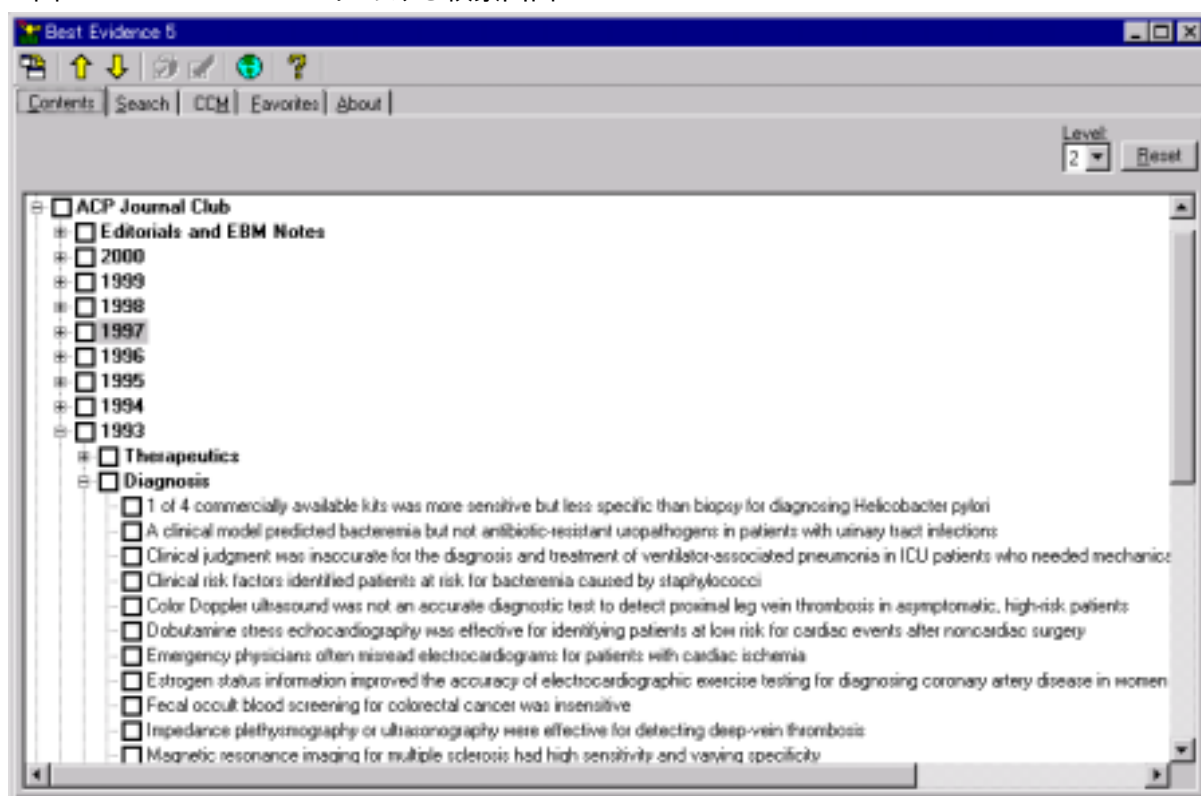


また、ACP Journal Club と Evidence-Based Medicine の両雑誌を網羅した Best Evidence5(2001)では、元々、Diagnosis といった分類項目があり、その件数を検索した。

[†] Medline の検索と重複する可能性が高いが、質の高い review のみが掲載されている。

その結果、ACP Journal Club と Evidence-Based Medicine を加えた記事総数は、1817 でそのうち、Diagnosis に分類されたのは、158 であったが、その他の etiology や prognosis などに含まれるものも入れると、同様に検体・生理機能検査に該当したものは 164 であった。

図 5. Best Evidence 5 における検索画面



-2. 検査領域における系統的再評価の実践における問題点について

-2-A 文献資料検索 - MEDLINE を対象として

系統的再評価(SR)において最初に行うことは、問題の定式化である。治療効果や病因などにおける SR と同様に、診断検査に関する SR においても、以下のような PECO あるいは PICO の形で定式化を行う。これによって臨床上的疑問が一層明確になり、再現性の高い evidence の検索が可能になる。また、定式化した問題は次のいずれかに分類しておく。

【分類】診断、治療、予後、頻度、副作用、予防、コスト

定式化の例：

問題：「近年、老人基本検診において男性受診者を対象とした PSA による前立腺癌のスクリーニング検査が話題になっている。果たして、PSA は従来行われてきた Acid Phosphatase に比較して、前立腺癌の診断に対する感度・特異度が優れているといえるであろうか。また、長期的に前立腺癌による死亡率を減少させるであろうか。」

【分類：診断】

P：Patient (対象)

老人基本検診を受ける男性

E：Exposure(暴露) あるいは I：Intervention (介入)

PSA によるスクリーニング

C：Comparison(比較)

Acid Phosphatase によるスクリーニング

O：Outcome (アウトカム・帰結)

前立腺癌の診断に対する感度・特異度

前立腺癌による死亡率の変化

文献検索においては、このように定式化された内容を主なキーワードとして検索を行うことになるが、ベッドサイドの EBM では一般に以下のような情報源を検索の対象とする。

1. 教科書
2. ガイドライン
3. 2次峻別文献集あるいは EBM 関連ジャーナル
The Cochrane Library (Update software)
Best Evidence (ACP/ASIM)
4. 文献データベース
MEDLINE
(Index Medicus)
EMBASE
HealthSTAR
医学中央雑誌

JMEDCINE

JAPICDOC

日常的な臨床における EBM の実践では、上記 1～4 の順序で目の前の患者に適用できる evidence を検索することになるが⁴⁾、SR を行う場合には、二次的 evidence を創ることを目的とするため、4 の文献データベースから関連した論文が抜け落ちないように系統的・網羅的な検索ストラテジーを立てる必要がある。

そこで、文献情報の検索時におこる問題について、MEDLINE を対象に検討した。

-2-A-1. 文献検索の手法

検討にあたり、まず複数の課題を提示し、それらに対する回答 (evidence) を得るために各班員 (A～E) が実際に検索を行い、それぞれの検索結果を持ち寄り比較した。例えば、「HbA1c 測定における HPLC と Immunoassay との乖離はどの程度問題になるのか」という課題に対し、表 1 (Q2) のような検索結果を得た。

検索のためのプロバイダは、A: SPIRS (Silver Platter 社), B: MD consult ([Harcourt General, Inc.](#)), C～E: Ovid (Ovid Technologies 社) である。検索に用いたキーワードは、1) HbA1c, Hemoglobin-A-Glycosylated/analysis, 2) Immunoassay, 3) HPLC, Chromatography, High Pressure Liquid であり、班員により得られた最小と最大の文献数はそれぞれ 1) 2,891～9,364 件 (表 1 #1) 2) 27,977～210,410 件 (#2) 3) 38,567～90,789 件 (#3 or #4) であった。これらを最終的に論理積 (AND) で結んだ結果 (#7) は、A: 42 件, B: 50 件, C: 27 件, D: 17 件, E: 46 件 であった。

同一の課題に対して検索により得られた文献数にこのような差を生じた原因として、以下のことが考えられた。

- (1) プロバイダによる相違
- (2) キーワードによる相違
- (3) MeSH Terms と Text Words の使用による相違
- (4) subheading の指定と explode の使用による相違

(1) プロバイダによる相違

インターネットを介し MEDLINE にアクセスするための各種プロバイダが存在するが、それらの検索エンジンの違いなどから、同一のキーワードを用いても得られる文献数には大小さまざまな差を生ずる。その検索の仕組みの違いはユーザーにはブラックボックスとなっている。また、大学や医療施設によって文献情報を入手 (電子検索) できる環境も異なっている。因みに Ovid 社の MEDLINE を導入している大学は、全国の 80 医学部のうち 60 ぐらいといわれる。

(2) キーワードによる相違

選択されるキーワードは、課題に用いられた用語に依存しがちである。したがって、課

エクセル表 1 が入ります。

題の提示(すなわち問題の定式化)に際しては適切な用語の使用が望ましい。検査に固有の用語については、複数存在する同義語によって検索結果も異なってくることに留意する必要がある。

(3) MeSH Terms と Text Words の使用による相違

MeSH (Medical Subject Headings)とは、夥しい数の医学用語を、それぞれに派生する類義語とともにツリー構造に編集した辞書 thesaurus をいう。National Library of Medicine (NLM)の indexer (インデックスをつける人)はこれを用いて、各論文にキーワードを与えている。MeSH には、検査に関連した用語がすべてインデックスされている訳ではない。検討した課題では、MeSH 用語として "Hemoglobin A, Glycosylated" が存在したが、"HbA1c" は存在しなかった。

一方、Text Words は、論文のタイトルや抄録のテキストデータに用いられたすべての用語を対象としている。Ovid の場合、mapping (.....mp.) によって、指定した Text Words をタイトルおよび抄録に含む、すべての文献を得ることができる。MeSH 用語に存在しない "HbA1c" の場合、Text Word として mapping 検索すると良い。

また、通常は1つの論文に複数の MeSH 用語が割り当てられるが、indexer は必ずしも検査を専門とする訳ではないので、常に最適な MeSH 用語がキーワードとして登録されているとは限らない。したがって、MeSH 用語による検索のみで目的とする文献のすべてが得られる訳ではない。このため網羅的な検索には、MeSH 用語による検索に Text Words による検索を併用することが望ましい。

(4) subheading の指定と explode の使用による相違

subheading とは、MeSH 用語で扱われる特定のフィールドをいう(Hemoglobin-A-Glycosylated/ analysis の "analysis" に相当)。MeSH 用語で検索する際に subheading を指定することで、目的に叶った論文に絞り込むことが可能であるが、前述の indexer の問題から抜け落ちる論文も少なからずあると考えられることより、SR のための網羅的検索には不向きである。したがって、MeSH 用語には subheading を指定せず、all subheadings として検索する方が良い (例: Hemoglobin-A-Glycosylated/ all subheadings)。

explode とは、MeSH 用語のツリー構造の中から、より下位に属する用語のすべてを用いて検索する機能である。キーワードとして用いた MeSH 用語がツリー構造のより上位に位置する場合に威力を発揮する (例: Immunoassay/ all subheadings 10,970 件 exp Immunoassay/ all subheadings 210,410 件)。網羅的な検索には MeSH 用語を explode させると良い。(SPIRS の場合、"Immunoassay"/ all subheadings と表記)

上記 2)~4)をまとめると、現時点で SR に最適な検索方法は、MeSH 用語について all subheadings を対象に explode 検索した文献 (exp [MeSH] / all subheadings) と、MeSH 用語に関連した考え得るあらゆる Text Words を用い mapping 検索した文献 (..... [TW].mp.), (.....[TW'].mp.), のすべてを論理和で結ぶ (or or or)やり方であろう。

-2-A-2. 文献検索上の問題点と今後の課題

-2-A-1 の検討結果をもとに、同じ課題に対して3種類のプロバイダで再度検索を行った。初回および再検索にて得られた文献数は、SPIRS: A(42件) A'(67件), MD consult: B(50件), Ovid: C(27件) C'(58件) (表2)であった。表3の共通文献数 - - は、それぞれ : A,B,C の3者に共通の文献数、 : 2者に共通の文献数、 : 単独の文献数、を表す。これらの検討から得られた知見は以下のとおりである。

(1) プロバイダの特徴

SPIRS や Ovid と異なり、MD consult は Text Words による mapping や explode といった操作を必要としないため、再検索による比較はできなかった。しかし、検討したいずれの課題においても、初回の検索で得られた文献数は3種類のプロバイダの中で最も多かった。さらに、他のプロバイダが Text Words を駆使して再検索すると、MD consult よりも多くの文献を得たが、それでも尚、MD consult でしか検索できない文献が少なからず存在した(表3)。これは同じMEDLINEを情報源としながらもMD consult 独自の検索エンジンによる違いと考えられる。検査用語に精通していない多忙な臨床医が、知りたいエッセンスを容易に抽出できるという点からすると、MD consult は使い易いプロバイダの一つといえるであろう。一方、SPIRS と Ovid の場合、同一のキーワードを用いて検索を行うとほぼ同一の文献が収集可能であった。

(2) キーワードの表記方法と問題点

表2A'に記載した“HbA1c”と“Hb A1c”、“HemoglobinA1c”と“Hemoglobin A1c”の違いは、いずれも Hb と A1c の間のスペース(半角)の有無である。両者を用いて得られた文献数には大きな差がみられた。SPIRS の場合、HbA1c: 2894件, Hb A1c: 155件, HemoglobinA1c: 808件, Hemoglobin A1c: 199件であった。また、“Haemoglobin” とするとスペースの有無にかかわらず、1つの文献しか拾えなかった。キーワードの選択のみならず、キーワードとなる用語の表記の仕方にも注意を払う必要がある。

(3) MeSH Terms と Text Words の併用効果と問題点

SPIRS による検索では、MeSH 用語の“exp Hemoglobin-A-Glycosylated / all subheading”で6,931件の文献を得たが、これに複数のText Wordsを併用すると8,659件となった。同様にして、Immunoassay [MeSH] が210,426 604,655件、Chromatography-High-Pressure-Liquid [MeSH] が75,961 92,926件へと増加した(表2, A A')。

Ovid の場合も、Text Words の併用で、Hemoglobin A, Glycosylated [MeSH] が6,923 8,214件、Immunoassay [MeSH] が210,410 217,282件、Chromatography, High Pressure Liquid [MeSH] が75,960 90,309件へと増加した(表2, C C')。このうち最も変化の大きかったChromatography, High Pressure Liquidを例にその内訳をみると、“exp Chromatography, High Pressure Liquid / all subheadings” [MeSH] と“HPLC.mp” [Text Words]で得られた文献はそれぞれ75,960件と38,567件で、最終的に両者を“or”で結んで90,309件の文献を得ている。

この数値の意味するところは、単に検索手法としてのText Wordsの併用効果だけでは

エクセル表 2 が入ります。

エクセル表 3 が入ります。

ない。MeSHにはない用語“HPLC”を用いなければ検索されない論文が14,338件(38,567件中, 37%)に上ったという点にも注意を払うべきであろう。このように論文中(とくにタイトルと抄録)の用語の用い方によって、検索上の分類も大きく変わり得るということは、キーワードを付与する indexer 側の問題ではあるが、その限界を考えた場合、論文作成の側にも何らかの対処の仕方があってよいのではないかと思われる。

一方、A'の例でも明らかのように、Text Words を数多く用いるほど得られる文献数は増加するが、操作の煩雑さに比例してその併用効果が必ずしも上がるわけではない。また、ある特定の Word でしか検索できない論文については、その数が明らかでなければ、当然ながら個々の Text Word による併用効果も不明である。

今後の課題として

- (1) EBLM/EBD に対応した検査用語集を作成する。また、それを独自のデータベースとして管理維持していくために、今後も増大する検査項目に備えて、タイムリーに対応できる登録システムを構築する。
- (2) 誰もが診断検査に関する一次、二次文献、その他の情報(精度管理や施設間差に関する情報など)を精度良く入手できるように、EBLM/EBD 独自のデータベースを構築するとともに(「EBD データベース」)インターネット等を介し文献検索する際に、これとリンクして適切な検索用語が選択可能となるような環境を整備する。
- (3) 診断検査に関する論文の検索手法を工夫するだけでなく、論文に使用すべき用語の選択や表記の仕方についても統一を図るなど、プライマリー論文作成の支援の一環として、用語に関するチェックシステムを導入する。

などが考えられる。

これらの課題を含めた文献検索ストラテジーは、SRのみならず日常的な EBLM/EBD の実践にも応用可能であろう。MEDLINE を対象とした今回の調査で明らかになった点は、いかにコンピュータを駆使しても、網羅的な検索には一定の限界があることである。それは MEDLINE のような膨大な情報を最終的に整理分類するのがコンピュータではなく“人”であるために生ずる限界である。それらを臨床検査医学の専門性で補完するようなシステムの構築も、EBLM/EBD の普及に向けた一つのステップであると思われる。

-2-B 論文の質の評価について

The Cochrane Library の Complete Review や Best Evidence などに収載された項目については、すでに批判的吟味(Critical appraisal)がなされたものであるので、内的妥当性についてはその Criteria をほぼ満たしているものと考えて良く、その Evidence を目の前に患者について適用できるかどうかの外的妥当性について論じれば良い。

しかし、SR の対象になる原著論文については、まず、その内的妥当性について検討す

る必要がある。無作為化症例対象比較試験(Randomized Control Trial : RCT)については、Chalmer T.C.などの質の評価法が提唱されてきた。⁵⁻⁷⁾ 診断検査についての論文の質については、Reid MC らが、彼らによって提唱された7つの基準を満たす論文は、1978年から1993年にかけては無く(資料2)⁸⁾、また、Lijmer らが1996年～1997年にかけてのSRに用いられた論文を調べたのものでは、わずかに6.8%が全てを満たしているといった結果であった⁹⁾ことから今後の研究デザインの改善のみならず、SRに際しては内的妥当性についての十分な批判的吟味が重要であることを指摘している。

診断検査におけるエビデンスレベルは、Centre for Evidence-Based Medicine (<http://cebm.jr2.ox.ac.uk/docs/levels.html>)では、5つのレベルに分けられるが、その中で、“independent”、“masked”、“consecutive patient”などの要件が満たされていない場合の結果への影響についても同様にLijmer らが報告している。⁹⁾(下表)

Study characteristic	Relative diagnostic odds	Description
Case-control	3.0 (2.0 ~ 4.5)	A group of patients already known to have the disease compared with a separate group of normal patients
Different reference tests	2.2 (1.5 ~ 3.3)	Different reference tests used for patients with and without the disease
Partial verification	1.0(0.8 ~ 1.3)	Most of positive results and a only small part of negative results are
Not blinded	1.3 (1.0 ~ 1.9)	Interpretation of test and reference is not blinded to outcomes
Non consecutive	0.9 (0.7 ~ 1.1)	Not all patients presenting the relevant condition are included into study.
No description test	1.7 (1.1 to 1.7)	Test not properly described
No description of population	1.4 (1.1 to 1.7)	Population under investigation not properly described
No description reference	0.7 (0.6 to 0.9)	Reference standard not properly described

表4 :The relative diagnostic odds ratio indicates the diagnostic performance of a test in studies failing to satisfy the methodological criterion relative to its performance in studies with the corresponding feature.

この結果より診断検査に関する論文の質のチェックは必須と言えるが、質の低い論文の取り扱いについてのコンセンサスは、まだ出来ていない。(Evidence and Diagnostics <http://www.jr2.ox.ac.uk/bandolier/booth/diagnos/Diagessy.html>)

今日まで、いくつかの診断検査に関する論文の質のチェックリストが提唱されているが、RCTにおけるチェックリストに比較すると、まだ開発段階にある。¹⁰⁻¹³⁾しかし、SR自体が、質的なクライテリアを満たさないものを除外する、あるいは、それらの混合から、差

異を明らかにすることであることから、不可欠な質的評価を行う際のチェックリストの作成を行った。その中で、今まで発表されている評価リストに共通なキーとなる確認項目は以下の通りである。

1. 標準的検査(Gold Standard, reference standard)と問題としている検査 (index test) の比較がそれぞれ独立して、マスク(ブラインド:Blind)をされた状態で行われているか？
特に、心電図などの波形による判定や形態学的な診断など判断に主観が入る可能性のあるものはマスキング(blinding)が不可欠である。
2. 問題としている検査の結果にかかわらず、標準的検査がなされているか？
倫理的な問題からより侵襲や危険性の高い標準的検査(Open biopsy など)は、その施行が難しい場合もあり、その結果、感度がより高く見積もられ、また特異度がより低く見積もられる系統的誤差(Test referral bias)が問題となる。
3. 検討された患者のスペクトラム(進行度・重症度・合併症など)が明確に記載されているか？
スペクトラムが異なる集団では感度、特異度が異なる。また、患者に適応する際にその差異を念頭に置く必要があり、必要な情報である。¹⁴⁾

加えて、Irwig L らが Cochrane Collaboration の Screening and Diagnostic Test Method Group(<http://www.cochrane.org/cochrane/sadt.htm>) でまとめた Recommended method (資料 1 にその日本語訳) や American Association of Clinical Chemistry の機関誌である Clinical Chemistry に Bruns DE らがまとめたチェックリストなどを参考に診断検査に関する論文のチェックリストを作成した。¹³⁾ 今回のチェックリストには、検査特性(analytical performance)について、その測定法や Coefficient variance (CV)などの明記を追加した。資料 2 に付記する。

このチェックリストは、SR における論文の質の評価のみならず、投稿論文についてのレビューの際にも参考となるものであり、ひいては、研究計画をデザインする際や投稿する際のチェックリストにもなるものである。このような試みはすでになされており、RCT 論文についての CONSORT 声明¹⁵⁾や SR やメタアナリシスの論文についての QUOROM 声明¹⁶⁾などが作成されている。以下に日本語訳がなされた Web サイトの URL を付記する。
(http://www.sphere.ad.jp/cont/CONSORT_Statement/menu.html)

一方、MEDLINE などの医学文献データベースを参照すると、ほとんどの原著論文に抄録が含まれており、SR における関連論文を検索する上で抄録の有用性と質が指摘されている。抄録には文献の内容が簡潔にしかも正確に記載されており、SR 時にはその内容から論文の採択・除外の際の重要な情報となる。最近ではこの抄録の記載法が、本文と同様に Introduction, Materials and Methods, Results, Discussion(IMRAD)といった構造化抄録の形で記載されることが多くなっている。¹⁷⁾

これは、Uniform Requirements for Manuscripts Submitted to Biomedical Journals といった統一投稿規格でも正式に推奨されており¹⁸⁾、その形式を採用する雑誌が増加して

いる。総合医学雑誌では、Annals of Internal Medicine, British Medical Journal, New England Journal of Medicine, JAMA など主要雑誌のほとんどが採用しているが、臨床検査関連の雑誌では、Clinical Chemistry や Archive Pathology and Laboratory Medicine など構造化抄録をとっているのに対して、Clin Chim Acta など従来形式の抄録を留まっているものも少なくない。

表 5 に臨床検査関連と一般の代表的な欧文雑誌の抄録の形式について調査した結果をまとめた。(2001 年 3 月)

Journal	構造化?	代表的フォーマット
Clinical Chemistry	○	Background, Methods, Results, Conclusions
Clinica Chimica Acta	×	
Clinical chemistry and laboratory medicine	×	
European journal of clinical chemistry and clinical biochemistry.	×	
Archive of Pathology and Laboratory Medicine	○	Objective, Methods, Results, Conclusions
Clinical and diagnostic laboratory immunology.	×	
American journal of clinical pathology.	×	
The American journal of medicine.	○	Purpose(Background), Methods, Results, Conclusions
Annals of Internal Medicine	○	Background, Objective, Setting, Design, Patients, Measurement, Results, Conclusions
ACP Journal Club	○	Objective, Setting, Design, Patients, Description of test and diagnostic standard, Main outcome measures, Main results, Conclusions
JAMA	○	Patient(Participant), Main outcome measure, Results, Conclusions
BMJ	○	Objective, Design, Setting, Subjects, Intervention, main outcome measure, Results, Conclusions
New England Journal of Medicine	○	Background, Methods, Results, Conclusions
CMAJ	○	Objective, Design, Setting, Patient, outcome measure, Results, Conclusions

表 5 : 代表的な論文と構造化抄録の採用 (2001 年 3 月現在)

構造化抄録の記載項目を臨床検査についての論文から考えると、objective, design, setting, patients and participants, intervention, measurement, results, conclusions などの項目が一般的であるが、診断検査の論文では検査方法や CV などのいわゆる analytical performance 事項を measurement の中に記載すべきと考えられる。これら方法の違いなどの情報は SR における研究間の層別化に活用できること、さらに目の前の患者に適用できるかどうかの判断に役立つものと考えられる。上記の雑誌の中でも ACP Journal Club が Description of test and diagnostic standard を個別の項目にしていることは特筆される。

-2-C. 統計的手法 (Meta-Analysis)について

診断検査における統計的手法についての提唱されている方法について調査した。調査した方法は、MEDLINE からの検索および参考文献に上げた書籍から行った。¹⁹⁻²¹⁾

【調査結果】

治療効果を RCT から検討する SR と同様に、用いる全て研究から統計的な推定(pooling estimation)や再現性を見るメタアナリシスは必須では無いが、検査の診断特性や危険予測における SR に際して行われることが多い。治療効果においては、通常、統計値(P)や Risk ratio, Odds ratio など一つの指標を対象にするのに対して、診断検査では、用いられる指標は、尤度比(Likelihood)やオッズ比(Odds Ratio)などの単一の指標が用いられることもあるが、通常は感度、特異度など2つの指標が用いられる。[‡]そのため、RCT におけるメタアナリシスとは異なるアプローチが必要となる。¹⁹⁾

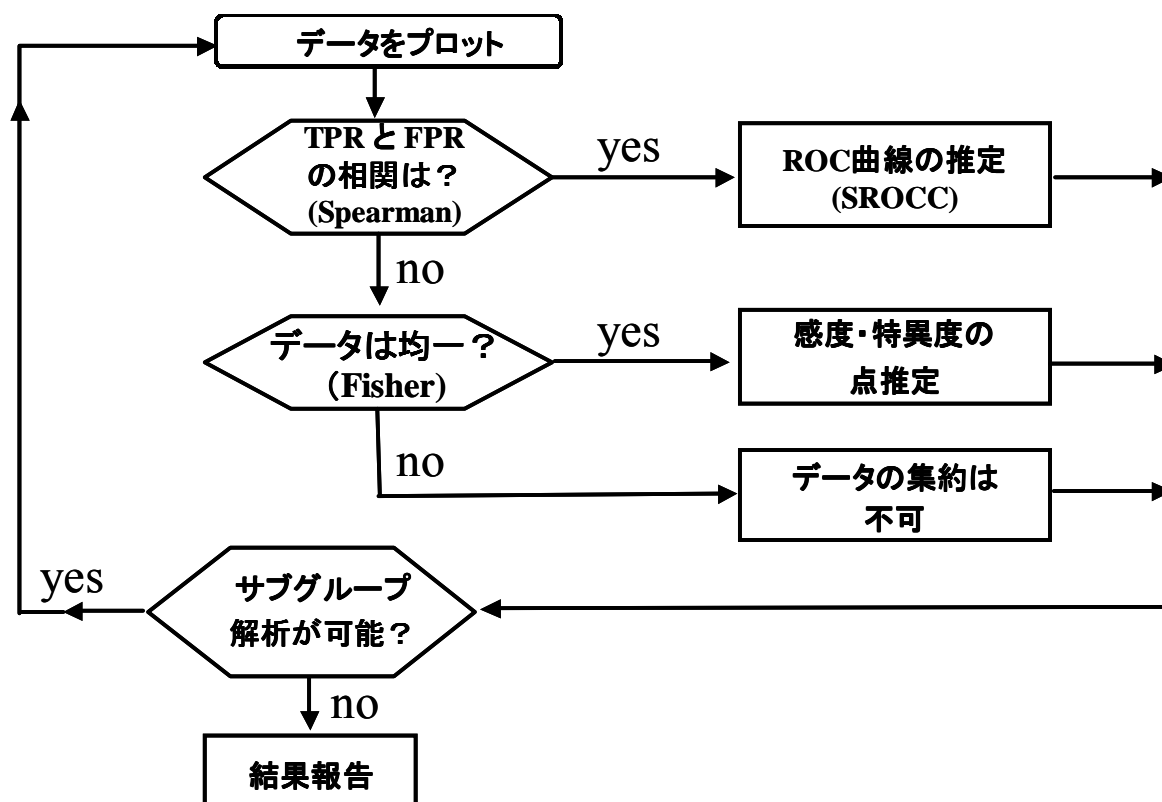
まず、研究間における診断能の均一性の評価であるが、RCT のメタアナリシスと同様、通常の偶発誤差(random error)以外に、患者スペクトラムの違いや研究デザインの違いなどの要因とともに陽性基準(閾値、カットオフ値)の違いによる影響を考慮しなければならない。多くの論文で単一の陽性基準による感度、特異度が示され、また、最適と考えられる陽性基準も疾患の有病率によって変わりうる。従って、感度、特異度をそれぞれ別個にそのサマリー値(Pooling sensitivity and specificity)として求める場合、₂ 検定などによる検定を行うとその多くが不均一となる。すなわち、閾値やそれにとりなう感度・特異度が大きく異ならないといった特殊な場合にしか適用できない。¹⁹⁾ このような背景から、現在、用いられているあるいは提唱されている方法は、次のようなものがあるが、₂ が使われることが一般的である。

1. 感度、特異度の点推定(加重平均による)
2. サマリー ROC 曲線 (Summary ROC curve)
3. 診断効果サイズ(Diagnostic Effect Size)

1,2 を組み合わせたアプローチのフローを Midgett らが提唱しており²²⁾、以下に示す。すなわち、それぞれの研究の診断特性である感度と偽陽性率(1 - 特異度)について散布図としてプロットし、その相関(Spearman)をみる。相関が見られる場合には、後述のサマリー ROC 曲線を推定することになる。この際に、感度、特異度の点推定(weighted average)を行うと、感度と特異度のトレードオフの関係を無視することになり過小評価になることが指摘されている。

次に、データが均一(homogeneous)かどうかの検定を行い、均一であれば、感度・特異度の点推定を行い、不均一であれば、その原因として見いだされた要因でサブグループに分類した後、それぞれのグループについて上記の手順を繰り返すものである。

[‡] Likelihood(LR)も陽性 LR=感度/(1-特異度)、陰性 LR=(1-感度)/特異度、オッズ比(OR)も OR=(感度×特異度)/[(1-感度)×(1-特異度)]と感度・特異度から導き出されるものである。

図 6 メタ分析のアルゴリズム例 (Midgette, et al ²²)

個々の方法について概略する。

1. Weighted average of sensitivity and specificity, likelihood, odds ratio and diagnostic effect size (感度、特異度、尤度比、診断効果サイズの点推定)

感度と特異度を別個に見た場合や尤度比をみるような場合に用いられるが、前述のように、研究間の陽性基準の違いによるバラツキが無い場合などに限り適用できるものであり、そのバラツキの有無を ROC 図にプロットするとともに、統計的(² 検定)に解析するなどあらかじめ検討しておく必要がある。

感度、特異度についての点推定の計算式は、以下の通りである。

$$p = \frac{\sum y_i}{\sum n_i} \quad SE(p) = \sqrt{\frac{P(1-P)}{\sum n_i}}$$

n_i は疾患のある患者(感度の場合)あるいは疾患のない患者(特異度の場合)の数

y_i は検査が陽性となった疾患を有する患者(感度の場合)あるいは検査が陰性となった疾患のない患者(特異度の場合)の数

尤度比の点推定については、RCT における fixed effect model あるいは random effect model による計算と同じであり、省略する。²⁰⁻²¹⁾

(http://ebd.umin.ac.jp/o_ef_size.htmlを参照)

2. サマリー ROC 曲線²³⁻²⁴⁾

陽性基準によって変化し、それがトレードオフの関係がある診断検査の感度・特異度をいわば独立した変数として取り扱う 1 の方法の限界から、この方法が使われることが多い。前提として、個々の研究の診断特性が異なるのは、(1)陽性基準のみによる、(2)(1)の陽性基準の差ばかりでなく、患者特性やその他の研究デザインの違いによるとする二つの仮定がある。²⁵⁾ (1)は fixed effect model であり、(2)は random effect model と呼ばれる。一般的には、fixed effect model である Littenberg B, Moses LE の方法が用いられる。²⁶⁻²⁷⁾

概略すると、検査が陽性の際の感度、特異度を組み入れた指標であるオッズ比は、真陽性のオッズと偽陽性のオッズとの比であり、以下のように示される。

$$OR = \frac{TPR / (1 - TPR)}{FPR / (1 - FPR)}$$

TPR : 真陽性率(感度) FPR : 偽陽性率 = 1 - 特異度

ここで、真陽性のオッズ比と偽陽性のオッズ比を自然対数に変換すると以下のように示されるが、

$$\text{logit}(TPR) = \ln[TPR / (1 - TPR)]$$

$$\text{logit}(FPR) = \ln[FPR / (1 - FPR)]$$

その上で次の 2 変数(D,S)を作成すると、

$$D = \text{logit}(TPR) - \text{logit}(FPR)$$

$$S = \text{logit}(TPR) + \text{logit}(FPR)$$

これらは線形回帰が可能となる。すなわち、 $D = \quad + S$ の関係となる。

ここで、非加重あるいは加重線形回帰分析やロバスト回帰式などパラメトリック、ノンパラメトリックの手法で、 \quad 、 \quad を算出する。ここで \quad は、陽性基準(閾値)による診断オッズ比への影響がなければ、オッズ比の自然対数を見ていることになる。 \quad は、そのオッズが陽性基準によって影響される度合いを示す。すなわち、 \quad が 0 であれば、ROC 曲線は感度=特異度の線について対称的な曲線になる。 \quad が 0 である場合に \quad は、陽性基準に関係なく一定であることから constant オッズ比あるいは common オッズ比と呼ばれる。(図 7 参照)

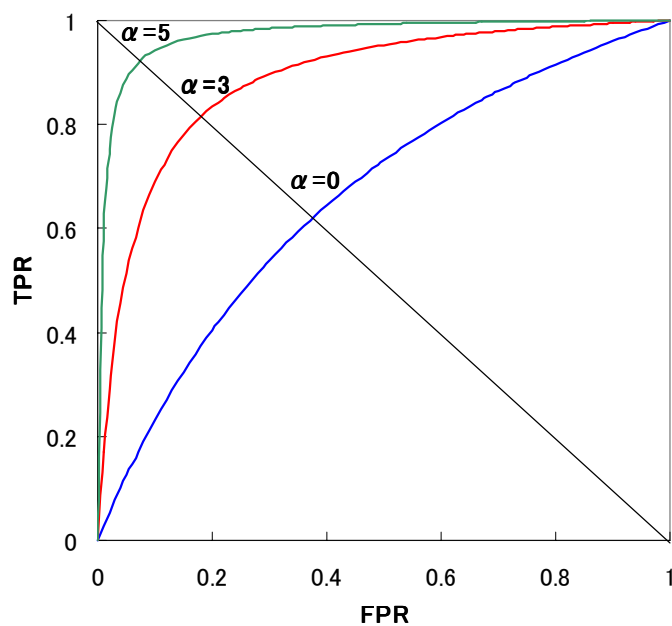


図 7: $\alpha=0$ 時の切片の影響

が大きくなるほど、判別能が高く、曲線が左上のコーナーに近づく。

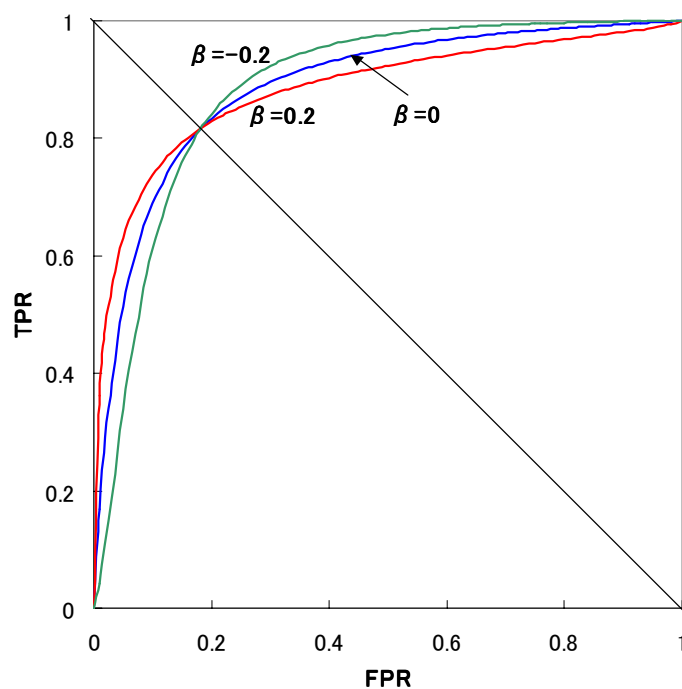


図 8 $\alpha=3$ 時の係数の影響

が 0 であると $TPR=1-FPR$ の線に対称となる。FPR 低値域にて、 > 0 では TPR が急峻な、 < 0 ではより緩やかな曲線となる。

Random effect model では、Latent-scale logistic regression model などがあるが、よ

り一般的なアプローチは未だに開発段階にある。²⁸⁾ また、Latent-scale logistic regression model によるメタアナリシスを行った論文は、検索した中では見いださなかった。(2001 年 3 月時点)

3. 診断効果サイズ²⁹⁾

疾患群と非疾患群との平均値の比較のように、2 群間の距離を標準化したものを物差しとして用い effect size とよぶ。特に複数の診断検査の比較に用いられる。

連続量の場合には[§]、

$$d = (M1 - M2) / s$$

$$s = \sqrt{\left[\frac{n1 + n2}{n1n2} + \frac{d^2}{2(n1 + n2)} \right]}$$

感度・特異度のように 2 値量、すなわち以下のような 2 × 2 表に結果が示されているような場合には**、

検査 \ 疾患		疾患	
		あり	なし
検査	陽性	A	B
	陰性	C	D
		n1	n2

$$d = \frac{\sqrt{3}}{\pi} \{ \log[Sn/(1-Sn)] + \log[Sp/(1-Sp)] \} = \frac{\sqrt{3}}{\pi} [\log(A) + \log(D) - \log(B) - \log(C)]$$

$$\text{var}(d) \approx \frac{3}{\pi^2} [1/A + 1/B + 1/C + 1/D]$$

として表される。

複数の研究から導き出された Effect size を Fixed effect model を使って統合する場合には、RCT と同様に、分散の逆数を重みとした inverse-variance weighting 式を用いて計算する。すなわち、

$$\hat{d} = \frac{(\sum w_i d_i)}{(\sum w_i)} \quad \text{Var}(d) = \frac{1}{\sum w_i}$$

$$w_i = \frac{1}{v_i}$$

[§] 疾患群と非疾患群が正規分布していると仮定した場合

** 分散が等しい場合

この effect size は、二群の平均の標準化距離であることから、非常にわかりやすいが、正規分布であること、分散が等しいことなどが条件であり、その仮説に従わないものについては、変形式が用いられるが省略する。

以上のように、診断検査についてのいくつかの統計的な手法が提唱されているが、以下のような問題が今後の課題として残されている。

1. heterogeneous な研究結果を解析する場合には、陽性基準の差異以外に患者スペクトラムの違いなどがある場合、後者がどの程度、診断特性値に影響しているかわからない、また、その場合、Random effect モデルなどの利用が必要であるが、その適用については言及されているのにとどまり、具体的な SR 例はなく確立されているとは言えない。従って、サマリー ROC 曲線は、最も一般的に用いられる手法であるが、診断検査の SR の標準とすることはできない。
2. サマリー ROC 曲線やコモンオッズ比などを求めても、それを直接、実際の臨床に当てはめることは危険で、対象とした集団の何らかの診断特性についての情報が必要であるためにその対象についての新たな実際の検討を行うことが必要と言ふことにもなりかねない。
3. メタ分析以前に質の低い、すなわち様々なタイプの bias が含まれる論文がどのように影響するか、また、その取り扱いをどうするかなどの経験が蓄積されていない。さらに、多くの論文で、どのような bias が含まれるかを判断するのに十分な情報が提供されていない。これら多くの要因が、現状での診断特性の SR についての限界につながっているとされ(Irwig L, Tosteson AN, Gatsonis CA et. al.: *Meta-analytical methods for diagnostic test accuracy. J. Clin Epidemiol* 48:119-130, 1995)、今後の発展が望まれる。

-2-D. 出版バイアスについて

少数例の検討や結果が有意差を認めないような negative な結果の研究が出版されずに SR の際に漏れてしまう系統的なバイアスを出版バイアスと言う。学位論文や政府機関や財団のレポートなども、検索対象のデータベースに含まれず漏れることは不可避と言える。このバイアスについては、有意な結果の論文の方が、有意でないものや標本数の少ない論文に比べ、オッズで 3.8 倍出版されやすいと言う結果もある。³⁰⁾ このような主として有意な結果が多く出版されている状況での SR は、その治療効果を過大評価し³¹⁾、また、副作用などが無視される危険性を持つことになる。

治療における出版バイアスについては、いくつか検討はなされているが、³²⁻³⁷⁾ 診断検査についての検討を行った論文は、今回の検討では見出すことはできなかった。また、RCT などの SR においては、出版バイアスを funnel plot といったグラフィカルな方法で検討し言及されることが多いが、診断検査の SR では、出版バイアスがより問題になりやすいと考えられるにも関わらず³⁾、明確な方法論を持って取り扱った論文はないとされる。³⁸⁾ また、RCT においては、IRB(施設内臨床研究審査委員会)などへの登録や、BMJ の Amnesty

for unpublished trials などの登録制度もある。

(<http://www.bmj.com/cgi/content/full/315/7109/622>)

しかし、診断検査の場合には日常診療のデータが用いられることが多く、出版されたもの以外の研究を見いだすことはまず不可能である。登録制度の試みも多くは成功していないようであり³⁹⁾、今後の大きな課題と考えられる。その中で、e-medical research として、主として Clinical Examination を対象とした共同研究(CARE study)が開始されている。⁴⁰⁾

(<http://www.carestudy.com/CareStudy/Default.asp>)

すなわち、インターネット上で研究参加者を募り、一定のプロトコールに従って、病歴や診察所見、検査結果と共にその health outcome を登録して貰うものである。このようなシステムを利用することが出版バイアスを克服する一方法となるものと思われ、今後の展開が期待される。

-2-E. Evidence の普遍性(Generalizability)とガイドライン

EBD の大きな問題として、SR によって得られた検査の診断特性について、臨床のそれぞれの現場にそのまま適用できるかという普遍性の問題がある。治療的介入における SR でも同様のことが指摘されている。例えば、人種の差では、冠動脈疾患では French paradox や Japanese paradox^{††}と呼ばれる発症率の低さが特筆されてきた。さらに日本人の冠動脈疾患では、高血圧や糖尿病などに比べ、高脂血症はそれほど大きな影響を与えていないと言った結果から⁴¹⁾、その対策は、アメリカとは異なったアプローチが必要とも言われる。

検査の診断特性についても、1論文で、組み合わせ検査において臨床設定(setting)に関係なく適応できるとしたものがあるが、同様の状況と考えられる。例えば、患者スペクトラムが違くと診断特性自体が異なってくることが挙げられる。すなわち、有病率には左右されないが、病気の人でも最も病的な人を対象にしたもの(Sickest of the sick)や健康人における最も健康な人を対象にしたもの(Wellest of the Well)では、診断特性が異なり、前者では、実際の臨床の患者より TPR(真陽性率：感度)、FPR(偽陽性率：1 - 特異度)が高くなりやすく、後者では、FPR が実際の臨床の場では上昇しやすくなることが指摘されている。

さらに、Ovid の検索システムにて、Medline、EMBASE および Cochrane Library、Best Evidence 5 のデータベースから下記の strategy で検索した。

Search History Results Display

1 Evidence-Based medicine.mp.	10429
2 limit 1 to diagnosis	10056

^{††} フランス人はアメリカ人と生活習慣や人種的にはあまり変わらず、高脂肪食、飲酒をとっているにも関わらず、冠動脈疾患の罹病率や死亡率が低いのは、赤ブドウ酒のポリフェノールによる抗酸化作用によると言われている。日本人では、アメリカ人に比べての高繊維食・低脂肪食や人種差がその要因と考えられている。

3 systematic review.mp.	4334
4 limit 3 to diagnosis	3949
5 meta analysis.mp.	23511
6 limit 5 to diagnosis	22401
7 applicability.mp.	22606
8 generalizability.mp.	2364
9 2 or 4 or 6	34388
10 7 or 8	24913
11 9 and 10	312

この中で目の前の患者への適用について論じた論文は、タイトルや抄録から関連するような文献でもわずかに 10 編程度であった。その 1 編は診断検査についての 23 の SR 論文を検討した報告であったが、ほとんどの研究でもととの研究デザインの違いや方法的な不備からプライマリデータそのものが不均一であることが多いとしている。従って、その heterogenous な結果からの統合値を臨床に用いること自体、問題となることは明らかと考えられる。⁴²⁾

以上から、眼前の患者と同じような集団についても適用できる普遍的な Evidence が望ましいが、検査特性におけるプライマリデータが heterogenous で普遍性が見られない場合には、SR によって研究デザインの違いなどのその原因となっている要因について解析し、明確に記載すると共に、眼前の患者の状況に近いデータが活用できるように関連研究について詳細に記載することが重要と考えられる。

また、今後、臨床で活用しうる Evidence を創るためには、日本での質の高い一次研究であるとともに自施設あるいは類似の医療機関での多症例のデータからのファクトを蓄積するが重要と考えられる。その基盤として日常の検査データを蓄積し、他の症状や徴候と言った診療情報と診断情報などと併せて二次利用できるシステムの構築が必要である。さらには、CARE STUDY のような多施設で共有するために検査の標準化とともに、共通フォーマットの作成、または登録システムなどの構築が望まれる。

一方、普遍性とは内容が異なるが、Evidence を基にしたガイドラインを作成することも Evidence へのアクセス性を高めるために必要である。一般臨床医の診断特性の指標(感度・特異度、尤度比、ROC 曲線)についての認識がまだ十分でないことから(表 6)、それらを直接的に示しても Evidence として活用されることが期待されないと考えられるからである。⁴³⁾

他方、診断検査はそれに引き続く治療効果にも関係してその有用性や費用効果などが論じられなければならないが⁴⁴⁾、最終的な治療効果を含めた Evidence は、Helicobacter Pylori と消化性潰瘍再発についてのものなどでまだ少ない。すなわち、最終的には疾患の診断とそれに基づく治療ストラテジーを含めた診療ガイドラインに組み込まれることが必要であり、今後、それらのもとになる一次的研究やそれに引き続く SR を行ってゆくストラテジーを検討すべきである。

表 6 Frequency of Reported Use of Individual Methods of Assessing Test Accuracy

Specialty	n	Bayesian Method	ROC Curves	Likelihood Ratios
Internal medicine specialty	50	5	1	1
General internal medicine	50	2	0	1
Pediatrics	50	1	1	0
General surgery	50	0	1	0
Family practice	50	0	0	0
Obstetrics/gynecology	50	0	0	0
Total	300	8(3%)	3(1%)	2(1%)

- 3 . EBD データベース

【 1 】 調査・検討経過

順天堂大学で試作したデータベース⁴⁵⁾を Base Model として、班会議おける情報交換、および “The Cochrane Library”, “Best Evidence 5”, その他インターネット上でアクセス可能な一次、二次データベースに関する調査、琉球大学医学部附属病院医療情報部 廣瀬康行 教授、名古屋大学医学部精神科 古川壽亮 教授からの情報教授等に基づいて、データベース構築の必要性、構造・仕様等について検討した。

【 2 】 EBD データベースの必要性と目的

EBM に基づく診療は、(1)臨床上的問題点の定式化、(2)問題点に関する情報の収集、(3)情報の批判的吟味、(4)患者への適応、という4つのステップを踏む。EBM を実践しようとする場合、(2),(3)のプロセスに多大な時間と労力が必要であり、MEDLINE のような一次データベースを適切に検索して批判的吟味を行うことは容易ではない。一方、教科書や “診療ガイドライン” は集約され吟味された情報を含むが、その多くは疾患を総論的取り扱うもので、焦点の絞られた個別の問題を解決する手法としての EBM には適合しにくい。したがって、EBM の概念を日常の診療行為に応用しようとするれば、個別の臨床的問題について “最新” で “最良” の情報があらかじめ集約され、それが利用しやすい形で存在し、かつ容易に入手できなければならない。

この目的のための二次的峻別情報源としては、“The Cochrane Library”, “Best Evidence”, “Clinical Evidence” などがある。しかし、いずれの情報源も治療効果に関する RCT などを主体としており、臨床検査領域については網羅的な情報源としての性格は有さない。また、現状ではインターネット上で入手可能な臨床検査に関する二次的情報源もほとんどない。したがって、evidence に基づいて検査を利用し、結果を解釈するという clinical practice としての evidence based diagnosis (EBD)を普及、支援するためには、臨床検査についての evidence を包括的に含む情報源 (EBD データベース) を構築し、インターネット等の媒体を介して容易にその情報を利用できるような環境の整備が必要となる。

この EBD データベースは二次的峻別情報源として機能することを目的とするものであるから、データベースから得られる情報(evidence)は信頼性が高く、最新の内容であることが保証される必要がある。したがって、データベースの作成にあたっては、現存する evidence を単純に収集・表示するのではなく、それを批判的に吟味し、必要に応じメタアナリシスを含む要約を行うという“evidence を創る”作業が不可欠になる。これは臨床検査専門家の collaboration で行うことになるが、その作業量が極めて大きい点が問題である。しかし、(人的資源をいかに確保するかという最大の問題を除き)最新のデータベース/Web 技術を利用することによって、この collaboration 作業自体もデータベースにリンクされた Web ページ上で効率よく行い、その結果をリアルタイムで出力に反映させるシステムを構築することが可能となっている。従来から存在する二次的情報源と本データベースとの差を明確にするという意味でも、このような情報技術の積極的な利用によって、evidence の伝達手段としてのみではなく、evidence の創出から利用までの EBLM/EBD 全体を総合的に支援できるデータベースを企画・作成することが望まれる。

【3】EBD データベースの構造と構築手法

(1) データベースの基本的構成

本データベース基本構造の一案を図9に示す。

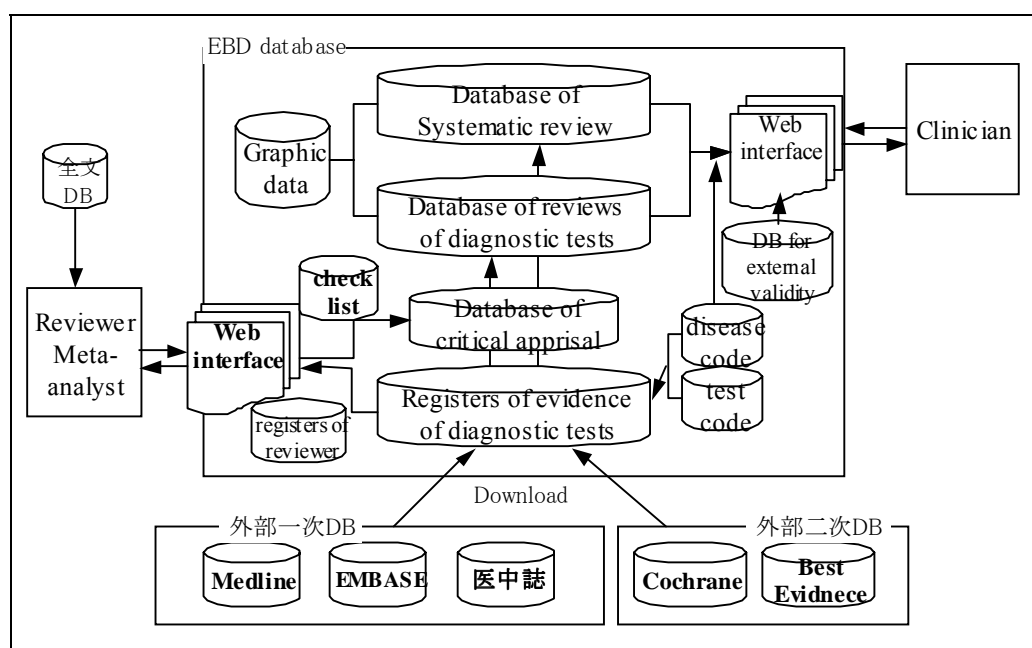


図9 EBD データベースの構成案

本データベース(DB)の中核は、“Registers of evidence of diagnostic test”とした table である。この table は外部の各種一次 DB、二次 DB から臨床検査関連の情報を網羅的に検索、ダウンロードし、アブストラクトを含む各フィールドを DB 化したものである。

本データベースに登録された Reviewer 向けには、その内容は Web interface (Microsoft

Active Server Pages)を介し、疾患、臨床検査毎に絞り込んで検索、表示を可能とする。Reviewer は論文要旨によるスクリーニングを Web Base に行い、全文を評価すべき論文についてはonlineまたはofflineにてこれを参照し、一定のチェックリストに従って、critical appraisal の結果と summary を入力する。この結果を保存するのが”database of critical appraisal”とした table である。

SR の原則に準じ、複数の reviewer にて一定以上の評価を与えられた一次資料は自動的に出力用(query) table である”Database of reviews of diagnostic tests”に追加される。また二次資料 (Systematic review) については同様に出力用の”Database of systematic review” table に追加される。Systematic review に関しては詳細な情報が得られるのが望ましいため、全文を網羅できる table 構造とする。有効な二次資料が存在せず、信頼性の高い複数の一次資料の成績に矛盾、バラツキが存在する場合はメタアナリシスを実際に行う必要があり、その結果はこの table に直接入力される。なお、Best Evidence (ACP journal club)に含まれる一次資料や Cochrane Library (CDSR/DARE)に含まれる二次資料については信頼性が確保されているものとして、上記の手順を経ず、直接出力用 table に追加するという方法も選択しうる。

一方、本データベースの利用者(主として臨床医を想定)は同じく Web interface を経由して、疾患別あるいは検査別に上記の2つの出力用テーブルを検索することにより summarize された診断特性を得ることになる。また同時にリンクを追うことによって個々の一次、二次資料の詳細(アブストラクト)を参照できる。また、その診断特性を利用するにあたって必要な外的妥当性についても DB 化し参照を可能とする。

(2) データベースの構築手法

1) 文献情報の取得

本データベースは臨床検査全般についての包括的情報源を目標とするため、現存する臨床検査関係の情報を網羅的(かつ効率的)に収集するという、通常の SR や臨床的問題の検索手法とは異なったアプローチが必要となる。

収集ストラテジーについては、MEDLINE を対象とする基礎的な検討を本研究班で行い網羅性の高い検索手法が提案されている。(-A 「検索ストラテジーのまとめ」)

本データベースは診断的検査の診断特性、およびその Systematic Review(SR), meta-analysis を対象としており、MEDLINE では以下のような用語による検索の組み合わせが、網羅性の高いデータ収集に向くと考えられる。

一次論文について

#1 (sensitivity- and-specificity) and (diagnosis)

#2 (sensitivity-and-specificity) and (diagnostic-procedures-and-techniques)

#3 (diagnosis) and {(risk) or (ROC-curve)}

二次論文について

#4 #1 or #2 and (Meta-analysis)

#5 #1 or #2 and (review)

検索結果は「 -1 検査医学における evidence の現状について」と重複するので省略するが、#1、#2 いずれも 10 万件を越える文献数であり、これをすべて一次データベース化するべきか、より特異度の高い検索手法を考慮すべきかは今後の課題である。しかし、Cochrane Library の CCTR/CENTRAL における”Diagnostic Techniques and procedures”をキーワードとする登録論文数が 48,000 件余り存在することから、大幅な特異度の改善は望めない可能性が高い。(EMBASE については未検討)

一方、国内文献は「医学中央雑誌(CD 版)」を対象に検索を行った。医学中央雑誌は”MeSH”に準じた日本語検索インデックスを有しているが、MEDLINE では 20 万件を越えるヒット数となる”sensitivity and specificity”に対応する検索用語、”感受性と特異性”によるインデックス検索では全年度分(1987～2000)でも 100 件以下の文献しか得られない。これは医学中央雑誌の indexer が適切なインデックスを付与できていないことを意味している。

MEDLINE と医学中央雑誌の両者に存在する論文を例に、両者のインデックスの差を以下の論文で比較した。[Maeda H et al, “前立腺癌発見のための有離対総 PSA 比と PSA density の価値”(医学中央雑誌収録のまま；原題は英語) Int. J. Urol. 1998 Jul;5(4):343-8] 以下が結果である。

MEDLINE (項目別)	医学中央雑誌
Aged / Aged, 80 and over / Middle Age Male Human Diagnosis, Differential Sensitivity and Specificity Predictive Value of Tests Survival Analysis Prostate-Specific Antigen/*blood Prostatic Hyperplasia/*diagnosis Prostatic Neoplasms/*diagnosis/mortality Support, Non-U.S. Gov't Japan	人/男 前立腺特異抗原(血液,測定・定量) 前立腺腫瘍(悪性,診断) 密度

上記から明らかなように、医学中央雑誌では”Diagnosis / 診断(鑑別診断)”, ”Sensitivity and Specificity (感受性と特異度)”など、本論文が検査による鑑別診断であることを特徴づけるキーワードが存在せず、実質的に本データベースに適合する検索が行えなかった。しかし、2000 年より運用が開始された WEB 版ではすべてのフィールドに対するテキスト検索が可能となったため、”診断/AL AND 感度/AL”, ”診断/AL AND 特異度/AL”の検索を行うことで件数は大幅に増加する。(それぞれ 3030 件, 772 件)

2) 検査項目および対象病態等の分類

上記でダウンロードされた一次、二次資料における検査項目、対象病態は、その後の検索キーとなることから適切な方法で Grouping を行う必要がある。MEDLINE, Cochrane, 医学中央雑誌は MeSH を用いており、他のデータベースの資料もこれに準じた Mapping

を行って一律に解析をできる環境を整備する必要がある。検査項目については JLAC10 への Mapping も考慮する。

3) 批判的吟味の定量的評価

本データベースでは一次資料の評価にあたって、巻末 資料4のチェックリストの(Web上での)利用を推奨したい。このチェックリストの各項目に対し、重要度に応じて適切な値を設定すれば、一次資料の信頼性の定量的評価を行うことができる。したがって、この定量的評価が一定以上であれば妥当性の高い資料としてユーザー側に表示する、あるいはSRの対象とする、といった対応が可能である。妥当な評価指標を今後検討していく必要がある。

(3) 臨床医側インターフェースの出力内容について

1) 臨床検査の evidence の捉え方

臨床検査の有用性に関する評価指標については、

- a) 検査の診断的利用における感度・特異度・尤度比を end point とする評価方法 (診断特性評価)^{§1}
- b) 検査とその結果に基づく医学的介入の効果^{§2}を end point として評価する方法
- c) a,b にコストの概念を加えた評価方法^{§3}

とがあり、いずれも臨床検査の evidence として捉えることができる。政策立案等に用いられる c は異質であるが、検査の実施によって患者予後がどれだけ改善されるかという意味では、医学的介入を含んだ評価である b)が望ましい側面もある。しかし、この評価法は一般に特定の診療モデルを前提とし、臨床検査自体の有効性に加え、対象疾患の予後、治療の有効性や副作用などさまざまな要因を反映するもので、検査自体の評価とは区別されるべき性格のものである。

前述のように、本データベースは EBM に基づく診療の実践における臨床検査の効率的利用と結果解釈のための情報源を意図しており、臨床検査自体の個別の診断特性評価を(主要な) evidence として提供することを基本とすべきであろう。加えて、効率的な検査利用を推進するという視点からは、同じ病態の診断を目的とする複数の検査が存在する場合、検査目的に合致する診断特性を有する検査を選択できるよう、一連の検査の診断特性評価を網羅的に表示・比較できる必要がある。

したがって、本データベースは“鑑別対象となる病態” “臨床検査” “診断特性” という形式で、病態別に臨床検査の診断特性を一覧・比較できる構成とし、必要に応じ検査項目別の(複数の病態に対する)診断特性や、一次資料の詳細な内容を閲覧できる機能

§1: 疾患の発症予測に関するコホート研究では relative risk (RR), odds ratio (OR)

§2: 研究デザインにより Utility や死亡率などの RR, OR, relative/absolute risk reduction (RRR/APR), number needed for to treat(NNT)などさまざま指標をとりうる。

§3: 費用便益分析(cost-benefit analysis), 費用効果分析(cost-effectiveness analysis), 費用効用分析(cost-utility analysis), 費用最小化分析(cost-minimization analysis)など

を有することが望ましいといえるだろう。

2) 診断特性の評価指標について

一次資料では、定性的検査では感度・特異度・尤度比(likelihood ratio; LR, 陽・陰性尤度比 LR+/LR-)が評価指標として用いられる。(但し疾患の発症・予後予測に関する検査の場合, RR/OR の利用も考慮する必要がある。) メタアナリシスを含む SR における診断特性は, SR の実施された時代や対象とした一次検討の内容により多様な分析手法が用いられている。SR の Outcome が感度, 特異度, LR+/- に summarize されている場合には, その情報は一次資料と同様に取り扱うことができる。

定量的検査(および多段階の結果を示す半定量的検査)では, 複数のカットオフ値を設定して感度・特異度を算出し, その変化を示す ROC (receiver operating characteristic) 曲線を用いるのが一般的である。ROC 曲線や area under curve (AUC)の比較は複数の診断的検査の比較に有効な手法であり, 検査施行前の有用性評価として複数の検査からより良質な検査を選択する場合に向く。したがって, 複数の検査比較についての一次資料の成績は ROC 曲線そのもの(および AUC)を表示する必要がある。これは, メタアナリシスの手法としてしばしば利用される summary ROC (sROC) 曲線分析でも同様である。

一方, 結果解釈に ROC 曲線から読みとった感度・特異度, 尤度比を利用すると, 有症患者, 無患者の測定値分布についての情報が失われるため, 誤った臨床判断に結びつくことがある。⁴⁶⁾ このため, 定量的検査については, 検討データに含まれる情報の欠落が少ない層別尤度比(Stratum-Specific likelihood ratio; SSLR)を評価指標として提示できることが望ましい。⁴⁷⁾

SSLR は測定値の層別に LR を求めたもので, ROC 曲線のスロープ(接線)に相当する。SSLR は一次検討における選択バイアスの影響を受けにくく, オッズ・尤度比法による検査後確率算出に直接利用できるという利点を有するため, 本データベースの主目的への適合性が高い。一方, SSLR の欠点としては, 測定値そのものを用いて層化するため測定値の外的妥当性の影響を受けやすいこと, 層化によって各層の対象数が減少するためカットオフ値による二分集計に比して信頼区間が広がる(したがって信頼性の高い SSLR を得るためには多数の症例が必要である)ことが上げられる。

3) 外的妥当性に関する情報

文献から得た evidence が個々の臨床的問題点に適応可能かどうか, すなわち外的妥当性の評価は本来情報の利用者側で判断すべき事項である。しかし, 臨床検査においては検査原理, 測定機器, 標準物質などの要因によりトレーサビリティ, コミュタビリティが未確立の項目も多く, 一般臨床医にとって外的妥当性の評価が困難である場合が少なくない。特に SSLR など測定値そのものを材料として判断を行う場合, 外的妥当性の誤解は臨床判断の過誤に直結する可能性が否定できない。

したがって, EBD データベースでは検査の診断特性とともに精度管理調査などに基づく

測定法間差，施設間差など外的妥当性を適切に評価するための情報を同時に表示できる必要がある。

(4) EBD データベースの確立に向けて

本データベースでは，情報内容の信頼性確保のために複数の reviewer による批判的吟味のプロセスを設定した。先にも述べたようにこのプロセスを実施する人的資源の確保が最大の問題であろう。このためには日本臨床検査医学会，日本臨床検査医会などの組織による強力な支援が必要である。また，できるだけ多くの検査関係者に collaboration に参加してもらう方策（例えば批判的吟味を行うことを認定医受験・更新資格とする，など）をとる必要があると考える。

-4. 自施設での Fact について

自施設で蓄積された検体検査のデータベースからもたらされた診断特性などは，普遍性の高い Evidence とは言えないが，その施設内での患者特性などを忠実に反映するものであり，その施設内での Evidence とみなして良いと考える。ここでは，一般的な Evidence とは区別して Fact と呼ぶことにする。

診断検査の Fact は，一般に日常検査データから派生するが，組み合わせて検討される病歴情報が retrospective なデータとなり，関連する情報に欠ける場合があるなど，一定の研究デザインのもとでなされたものとは情報の質において区別されるものである。一般的には，施設内での Case mix 的なデータと考えられる。そのため，病歴情報をもとに出来るだけ単一の病態の患者のみを選択して，その検査値をもとに尤度として表し，患者がそのデータから特定の病態としてのあるらしさ(類似度)を表すことが提唱されている。⁴⁶⁾ しかし，これは，主として診断特性の感度の部分を示しているのに過ぎないため，その検査値に影響するような他の病態が蓄積されて初めて，その特異度についても概略を知ることができることになる。さらに，比較的稀な疾患においては，施設での症例数に限界があるため，多施設でのデータを蓄積するシステムが望まれる。

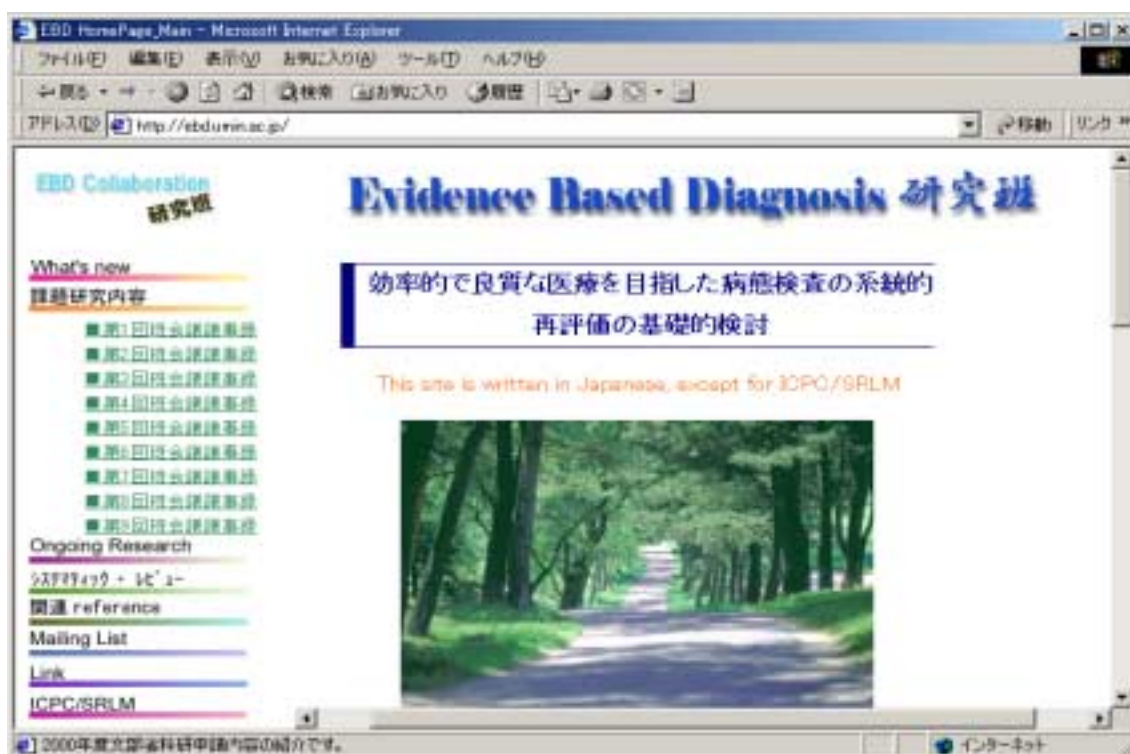
一方，多くの検査に関連した論文では，誌面の制限などから，結果は，あるカットオフ(判断)値における感度，特異度や 2×2 テーブルによる表示，あるいは，散布図や ROC 曲線図などで結果が示されているが，特に，結果が連続量で示される検査においては，このような表現(ものさし)を用いることで結果的に情報量が減少することになる。この情報量の減少を少なくする試みが前述の Stratum specific Likelihood ratio(SSLR)である。⁴⁹⁾ 施設内，あるいは多施設内で蓄積されたデータは，いわゆる生データであることから，情報の欠落を少なくすると言った目的から SSLR の適用は理にかなったものと考えられる。さらに，近似式や Fagan のノモグラムなどを作成することで，Computer 上での検査後確率の算出が可能である。[‡] この SSLR については名古屋市立大学精神科の古川敏明教授が，精神科領域の診断に適用して，その有用性を広く提唱している。⁴⁶⁾

[‡] Fagan のノモグラムは、<http://cebm.jr2.ox.ac.uk/frames/nomogram.html>にその例が示されている。

-5. EBD 啓蒙と協力体制について

日常臨床における EBM の実践では、できる限り短時間で効率よく必要な Evidence が得られることが望ましく、通常、教科書や二次峻別文献データベース(Cochrane あるいは Best Evidence など)などから、さらには、Medline から RCT など質の高い論文から Evidence を得ることが多い。しかし、SR では Evidence のもととなる一次データを網羅的に収集し、その上でいくつかのステップを介して Evidence を作成してゆくことになるが、厳密な論文の選択・除外や批判的吟味(critical appraisal)による論文の質の評価など、一連の作業は非常に時間がかかるものである。しかし、SR の課題となるような Evidence の明らかでない臨床的な疑問は、さまざまな領域に渡り、さらにこれまで伝統的に信じられ行われてきた医療の中にも Evidence の欠けるものが非常に多いのも事実である。このような背景から、The Cochrane Collaboration (<http://www.cochrane.org/>)などの国際的なボランティアベースの SR の組織が作られ、その中で分担し作業が進められている。また、重複した作業をなくすため、この中で登録制度も設けられている。一方、EBM についての様々な Workshop が開かれて啓蒙もなされている。The Cochrane Collaboration 主催の international colloquium の中でもさまざまな Workshop が開かれており、また、表にあげるようなさまざまな催しも予定されている。一方、国内でも CASP Japan(<http://www.phru.org/casp/>) や JANCOC(Japanese informal Network for the COchrane Collaboration)、愛知県臨床疫学研究会などが中心になって EBM セミナーが開かれており、その受講者の中に、SR を行える人材が育つことによって我が国の Evidence が創られてゆくことが期待される。

図 10 EBD の Top page



我々は、我が国における EBD の案内・啓蒙をはかる目的で、大学病院医療情報ネットワークを利用して、EBD の Web page を公開した。(図 10. <http://ebd.umin.ac.jp>) 今後、JANCOC や CASP Japan などとも連携して EBD の啓蒙を行い、多数が協力して SR による診断検査の evidence を創り、その蓄積をはかる必要があると考えられる。

- 6 . EBD をとりまく国内外の動向

【 1 】 調査経過

平成 12 年 4 月より翌年 2 月にかけて、(1) 関連ホームページ、(2) ICPC/EBLM 委員会、および(3) 以下に列記する学術集会・セミナー等(資料 6 参照)で得た情報をまとめた。

[1] 臨床 EBM 研究会第 5 回公開セミナーおよび CASP ワークショップ、2000 年 9 月 15 日(金)～同 16 日(土) 湘南国際村センター、神奈川県三浦郡葉山町

[2] 第 6 回アジア臨床病理学会、2000 年 10 月 11 日(木)～同 13 日(土) Haeundae (海雲台) Grand Hotel、韓国釜山市

[3] The 8th International Cochrane Colloquium、2000 年 10 月 25 日(火)～同 29 日(日) Good Hope Centre、ケープタウン市

[4] The 2nd China-Japan-Korea Joint Symposium on Medical Informatics、2000 年 12 月 2 日(土) Hyatt Regency Cheju、韓国済州島

[5] International Symposium on "EBM and Policy Development"、2001 年 2 月 16 日、都ホテル東京、東京

【 2 】 国内の動向(本研究班の活動を除く)

[ICPC/C-EBLM (Committee on Evidence-Based Laboratory Medicine, International Clinical Pathology Center)](<http://ebd.umin.ac.jp/eblm/>)

関連学会のトップクラスの関係者で構成され、事実上臨床検査医学界を束ねる EBLM 推進組織。実質的な活動は研究委託やセミナーの開催が中心であるが、委員長を務める河合忠国際臨床病理センター(ICPC)所長の豊富な人脈を活かし、ISO、IFCC、WASPaLM 等の国際団体、あるいは行政と密接に連携しながら活動している。

[日本臨床検査医学会 「日常初期診療における臨床検査の使い方」小委員会]

DRG/PPS やクリティカルパスへの対応を図るため、1989 年以降「日常初期診療における臨床検査の使い方」シリーズを刊行してきた同名委員会が、より臨床ニーズに則したガイドラインの確立を目指して発展的に解散し、1998 年に同名の新委員会として再発足した。その後本小委員会は、1999 年 4 月に 9 疾患群について「DRG/PPS 対応 臨床検査ガイドライン(第一次案)」を、次いで 2000 年 4 月に 14 疾患群を追加して第二次案を作成した。³⁾ 現時点では専門家の意見に基づいた記述が殆どであるが、将来的には EBM に対応した EBD のためのガイドラインを目指している。

[日本臨床検査医学会 EBLM 委員会]

学会として EBLM を推進する目的で 2000 年 12 月 16 日 (土) の理事会で設置が決まった委員会。この前身として、学術研究委員会の中の臨床検査情報学専門部に EBLM ワーキンググループが設置されていた。

[JANCOC (Japanese Informal Network for the Cochrane Collaboration)]

(<http://cochrane.umin.ac.jp/>)

日本におけるコクラン共同計画 (後述) の普及・啓蒙を主導する緩やかな人的ネットワークであるが、まだ Cochrane Centre の設立には至っていない。現在日本が参照すべき Cochrane Centre としては

「The Australasian Cochrane Centre (<http://www.cochrane.org.au/>)」が指定されている。

[CASP Japan] (<http://caspjp.umin.ac.jp/>)

CASP international (後述) の日本支部。CASP ワークショップの主催を中心に活発に活動している。今のところボランティアベースでの運営であり、人的ネットワークの拡充による作業負担の分散が課題である。

[臨床 EBM 研究会]

EBM 公開セミナーを年 1 ~ 2 回程度、これまで 5 回に渡り主催しているゆるやかな組織。

[その他]

EBM に取り組んでいる組織はこの他にも数多く存在すると推測されるが、調査した範囲では、EBLM に的を絞って活動しているものは上記の他に見当たらなかった。

【 3 】 国外の動向

[The Cochrane Collaboration] (<http://www.cochrane.org/>)

医療や健康に関するシステマティックレビューの作成、改訂、普及を目的としたボランティアベースのゆるやかな組織。運営委員会、共同レビューグループ、方法論グループ、コンシューマ・ネットワークなどから構成され、世界十数ヶ所のコクランセンターを拠点として、Cochrane Library の発行、教育研修、学術集会の開催等を行っている。

The 8th International Cochrane Colloquium で出席したいくつかのセッションでは、資金面での質疑応答が目につき、補助金を得た事業は部分的に盛んになるが、それが途絶えるとじり貧になるという、ボランティアベースの運営に共通する問題点が表面化しつつあるように感じられた。

EBD 関連については、Screening and Diagnostic Tests という方法論グループが 1996 年に SRLM (Systematic Review in Laboratory Medicine) のガイドラインチェックリストを発表しているが、共同レビューグループについては、方法論の確立が不十分という理由でまだ発足していない。現在方法論グループでは関連する他のデータベースやプロジェクトの動向などを検討しているが、マンパワー不足から、ガイドラインの改訂はおろか、ウェブページの更新、解析ソフトの充実、他組織からの問合せへの対応等にも事欠く状態

で、当面精力的な活動は期待できない。ただし、Likelihood Ratio (LR) のデータベースを作ろうという Bayes Collaboration の提案があり、その関連で紹介された Mike Pignone 氏 (University of North Carolina) が作成中の臨床検査項目の LR リスト (<http://www.med.unc.edu/medicine/edursrc/lrmain.htm>) は注目される。

[CASP international] (<http://www.phru.org/caspinternational/>)

EBM 的観点から、医療や健康上の意思決定の根拠となる情報を批判的に吟味する技能を、実践的ワークショップを通じて一般市民に身に付けさせることを目的とした、CASP (Critical Appraisal Skills Programme) という運動を進めている国際団体。

[IFCC/C-EBLM (Committee on Evidence-Based Laboratory Medicine, International Federation Of Clinical Chemistry And Laboratory Medicine)]

(http://www.ifcc.org/divisions/CPD/handbook/hdbk2000_2002/EMD2000.htm#Evidence Based)

IFCC が臨床検査医学におけるシステマティックレビューの実践を目的に立ち上げた委員会、現在は方法論の研究、他団体の動向調査あるいは他プロジェクトとの連携模索などを行っている。活動方針のひとつに、STARD (Standards for Reporting on Diagnostic Accuracy) という興味深い名称のプロジェクトがリストアップされているが、昨年9月にアムステルダムで会合があったこと以外詳細は不明である。

[その他]

EBM に取り組んでいる組織はこの他にも数多く存在すると推測されるが、調査した範囲では、EBLM に的を絞って活動しているものは上記の他に見当たらなかった。なお我が国の EBLM の動向については、特に韓国が強い関心を寄せている。

【4】EBD 研究における本研究班の位置づけ

EBM については、コクラン共同計画を中心に様々な活動が展開されているが、EBD は介入試験とは異なる方法論が必要となるため、国内外を通じてそれを模索している段階である。本研究班も独自の発想から文献評価のガイドラインや EBD データベースの開発を進めており、それらの一角を担う動きとして位置づけられる。

国内には EBLM の推進を目的とした組織がいくつか作られているが、最も実質的な研究を組織的に進めているのは本研究班である。したがって今後は、これまでの方法論の開発に加えて、膨大な実務作業を伴う EBD の実践を、国内および国外の研究者と協力し、どのように進めたらよいか、という課題にも取り組む必要があると考えられる。

なお、ボランティア的な熱意と努力に頼って研究を進める場合には、組織の運営負担が研究そのものを圧迫しないよう、コクラン共同計画を他山の石として、何らかの工夫が必要と思われる。

-7. 今後の課題

Evidence に基づいた臨床検査医学の今後の発展を考えるにあたって、大きく分け3つの立場がある様に思われる。第一の立場は、evidence の作成に寄与することである。すな

わちプライマリー論文の作成である。さらに質の高い論文が備えておくべき条件を明示し、質の高いプライマリー論文の作成を支援することもこの範疇に含まれる。

第二の立場は、臨床検査医学関連の evidence の集積である。すなわち、データベースの作成である。MEDLINE のような汎用データベースにはない特徴を持たせた臨床検査に特化したデータベースが作成されればこの分野に寄与することは大きいと思われる。

第三の立場は evidence の利用の場面において、その有効性を高めることである。臨床医、検査技師に直接かかわる点であり重要である。これらの狭間の問題、あるいは2つ以上の範疇に関わる問題もあろうが、あえてそれぞれに分けて今後の課題を論じる。

1. evidence の作成とその支援

evidence の作成は、すなわちプライマリー論文の作成そのものであり、それは全ての臨床研究を包括する。研究対象や研究内容はさておき、evidence 作成の立場から、全ての研究者に evidence として利用される質の高い論文が備えておくべき条件、あるいはチェックリストを示すことが最重要であると考え。この条件は、この研究班が中心となって早い段階に公表されるべきと考える。すなわち、これらの条件とは、本格的研究のみならず、従来の機器や試薬の検討の論文においても、感度・特異度や LR・SSLR とそれを導いた対象とデータなどを明記するなどすぐにでも出来る内容とおもわれる。検査の領域では多くの英語、日本語の原著論文が公表され、たとえ各々が十分な対象サイズでなくともこれらの条件を満たせば将来メタ分析の対象とすることが出来ることなどを周知し、全ての臨床研究が将来の evidence として利用できるものとなるよう研究者や雑誌編集者に啓蒙する必要がある。

また、種々の問題を含んでいるが、日常の検査結果の集積を自施設の fact として明らかにし、一つの evidence として扱おうとの試みもなされている。この手法の長所、短所を明らかにし、さらに標準化した手法の提案などもこの研究班の今後の課題である。

2 次的 evidence の作成も重要な今後の課題と考える。すなわち、メタ分析による 2 次的 evidence の作成や evidence に基づくガイドラインの作成などである。これは後に述べる evidence の集積や、evidence の有効利用と密接に関連する内容である。ここでは、この研究班の活動としてこのような 2 次的 evidence を作成する手法の標準化を目指すべきであることを指摘しておく。

evidence に基づく臨床検査を目指す上で、全ての研究論文が evidence として充分活用できる内容であることが本質的に重要なことであり、また理想像であることを銘記し、これの実現がこの研究班の今後の課題の大きな部分を占めるべきであると考え。

2. evidence の集積

臨床検査の分野で evidence に基づいた医療を目指すならば、関連する論文を過不足なく集積し、さらにそれらのプライマリー論文を適切に評価・吟味したデータを含めたデータベースの構築が望まれる。

このデータベースの内容や構築の手法に関しては色々なものが考えられそれぞれに長所

と短所があるであろう。日本語論文を含めた medline のようなものを目指すのは大変な労力と経費がかかり、現実的ではない。また、Cochrane Library のような 2 次的なデータベースは適切に評価・吟味されたもののみを集積しており、利用者にとって極めて有用と考えられる。しかし、このような 2 次的なデータベースの構築であっても大きな労力と経費がかさむことが考えられ、また維持していくことの困難さも想像される。

この研究班が臨床検査関連の論文の系統的再評価を目指すならば、系統的再評価をしたものの公表の手段、すなわち研究班の活動の公表の手段として再評価論文のデータベースの構築は避けて通れない問題である。利用者のニーズ、利便性などを考慮し、他のデータベースにない特徴を持たせたデータベースとはどのようなものか、そしてその構築、維持システムはどのようにするか、研究班員や協力者でより深い議論をする必要がある。目標とするデータベースの内容、サイズ、特徴などが固まり、維持体制に見通しができたならば早急にプロトタイプを構築するなど現実化することが今後の課題である。

3. evidence の有効利用

流行とも言える evidence based medicine の浸透により、medline を初め多くのデータベースが整備され、evidence に基づいた臨床検査を实践する環境も良くなった。しかし、実際に evidence に基づいた臨床検査が行われているかどうかに関しては、否と答えざるを得ない。その理由として、1) 一般臨床医、臨床検査技師に evidence に基づいた臨床検査の概念や必要性が充分浸透していないこと、2) evidence を求めるハードが整備されても、ソフトの面でのトレーニングがなされていないこと、3) 検査部をはじめとする検査施設はオーダーされた検査項目について正確な検査結果を出すだけが業務と捉えられ、オーダー際しての適切性の吟味、検査結果の臨床的意義の説明などについて考慮する機会が少ないことなどがあると思われる。

この研究班の今後の課題の大きなものとして、evidence に基づく臨床検査の重要性を、その概念や、実践の方法とともに広く啓蒙する必要がある。具体的には、第一に医師や技師の卒前教育において、evidence に基づく臨床検査をカリキュラムに取り入れる働きかける。その際には具体的なカリキュラムの内容や教材の提供が必要になる。この研究班の活動として、カリキュラムの作成と、教材・教科書の作成を提案したい。第二は、卒後教育の場面での啓蒙活動である。検査医学会、検査医会、技師会などと協力し、機会あるごとに啓蒙活動を行う必要がある。これらに関しても、研究班が中心となり、シンポジウム、ワークショップの企画、教材の作成などを行うべきである。EBM が浸透した経緯に、ワークショップや勉強会が積極的に行われたことを忘れてはならない。第三にはホームページ(HP)の活用があるであろう。研究班の活動の一部として HP に啓蒙的、教育的内容を充実させるなどの手法が考えられる。

なにごとにおいても社会に対する啓蒙活動は、地道な活動を不断に行うことが重要であり、啓蒙活動を研究班の今後の活動の柱のひとつとすべきと考える。

evidence に基づく臨床検査を広く実践していくと、例えば全ての検査項目において、基準

値を LR 値や SSLR 値に置き換えることや、全ての疾患において evidence に基づく診断ガイドラインや診断フローチャート、診断アルゴリズムを提案するなど現在の臨床検査医学を大きく改革することも可能となる。小さな研究班の活動としてはあまりにも大きな変革であり活動目標とは出来ないが、投げられた一石が水面に大きな波紋を広げる様に、この研究班の今後の活動が大きな変革の引き金になるようにしたい。

・ 総括

日常的に用いられている病態診断検査についての系統的再評価を経た Evidence はまだまだ少なく、そのため、重複して類似の検査や、診断精確度の低い検査が多数なされている状況にある。今後の医療資源の有効活用やクリティカルパスといった中での検査の効率的な活用に当たって、診断検査の精確度は大変重要であることは明白であり、従来よりも厳密な情報が要求されると考えられる。

今回の調査では、以下のことが明らかとなった。

病態検査の Evidence は、Medline や二次峻別データベースである The Cochrane Library や Best Evidence 5 をした結果、まだまだ少なく、治療に関するものに比べてその率も低い状況であった。

SR においては、関連する論文が抜け落ちないような網羅的な検索が必要となるが、検索ストラテジーによって検索結果は異なる。今回の検索結果から、MeSH Term と Text words の使い方や subheading や explode などの Filter 機能の活用に工夫がいること、プロバイダの違いも関係することが明らかにされた。

論文の質の評価は、論文の内的妥当性を検討するものであり、いくつかのキーとなる確認項目が共通に提唱されている。それぞれ、詳細部分で異なるが、提唱されたチェックリストの本質的な違いは無くなって収束しつつあると考えられた。

チェックリスト(methodological standard)から見た検査診断関連の一次論文は、ほとんどが研究デザインの不備によってバイアスの多いものになっており、そのため、RCT における CONSORT 声明のようなチェックリストを論文投稿時に活用することが望まれる。

近年、多くの雑誌に取り入れられている構造化抄録は、論文の概要が明確に記載されるシステムであり、関連する論文を取捨選択する際に、論文の内容を素早く吟味できることから有用と考えられる。

診断検査のメタアナリシスは、感度、特異度の加重平均やサマリー ROC 曲線、診断効果サイズなどの指標で行われるが、判断値のみが異なることを前提とした fixed effect model とその他の患者スペクトラムなどの背景因子が異なることなども考慮に入れた random effect model がある。それぞれの活用の仕方に一定の指針は未だに確立されておらず、また、random effect model

をベースとしたサマリー ROC 曲線による SR の具体例はまだ無い。
 診断検査の統計学的指標には、サマリー ROC 曲線や尤度比、オッズ比などがあるが、さまざまな要因から研究間で診断特性に違いが生じている場合が多く、その対策が確立されているとは言えず、その対策なしに統計的にまとめることは批判の多いところでもある。
 診断検査における出版バイアスの SR への影響は、治療の RCT に比べて検出することが難しく、また、funnel プロットのようなグラフィカルな方法も存在しない。そのため、CARE Study のような登録制度などが考慮される。
 SR の外的妥当性(generalizability)について言及している論文は少なく、また、検査の SR についての調査をしたほとんどの論文で、プライマリデータの研究デザインが不備であり、従ってその結果も不均一であることが多く、その pooling 値を臨床現場にそのまま適応することには慎重でなければならない。

これらの問題を克服してゆくために、以下のような今後の課題が挙げられた。

一次研究論文のプライマリデータの質を上げるために、研究デザインからバイアスを少なくするための診断検査における論文のチェックリストを導入し、広くコンセンサスを得ることが必要である。

二次峻別情報としての SR 結果へのアクセス性を高めるために、それらの結果をデータベースに蓄積し Web(HP)でのアクセスを可能とすべきである。また、その中でより最新の Evidence を維持するために、必要に応じて、臨床検査専門医が協力して Evidence を作るという作業環境の提供を可能とすべきである。

SR による Evidence ばかりでなく、自施設での Fact の蓄積も行うことで、外的妥当性をたかめ、さらに、将来的に多くの施設でのデータをメタ解析的に集約することで、Evidence のレベルを高めてゆくことも必要である。そのために、集積するデータの要素を過去の論文から SR によって明らかにすることや診断基準の統一などの方法論を検討すべきである。

臨床検査医が協力することが不可欠であるが、まだまだ、EBD についての理解が十分でないため、外部的なセミナーへの参加や様々な催しの中でその理解や実際のスキルの向上を図る必要がある。

Cochrane Collaboration など海外の団体とも連携を取りながら、協調して診断検査の SR の蓄積に邁進すべきである。

診断検査の最新で最良の Evidence が存在して初めて、質の高い診療ガイドラインの作成が可能となるものであることから、診断検査の Evidence を創り、更新し、さらにアクセス性を高めて臨床に活用して貰うと言った Cochrane 共同計画の挙げているスローガンと同様に、有志が診断検査の分野で共同して Evidence の蓄積を行って行くことが必要で

ある。そのためには前述したような多くのハードルが存在するが、臨床検査医が中心となって前進して行くことが重要と考えられる。

・ 謝辞

今回の調査にあたり、琉球大学医療情報部、廣瀬康行教授、名古屋市立大学精神科医学、古川壽亮教授、京都大学付属病院総合診療部、野口善令先生に、大変貴重な情報、御提言等を戴きました。深く御礼を申し上げます。

V. 資料

資料 1

Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001)

	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Differential diagnosis/symptom prevalence study	Economic and decision analyses
1a	SR (with homogeneity*) of RCTs	SR (with homogeneity*) of inception cohort studies; CDR† validated in different populations	SR (with homogeneity*) of Level 1 diagnostic studies; CDR† with 1b studies from different clinical centres	SR (with homogeneity*) of prospective cohort studies	SR (with homogeneity*) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval‡)	Individual inception cohort study with > 80% follow-up; CDR† validated in a single population	Validating** cohort study with good††† reference standards; or CDR† tested within one clinical centre	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way
1c	All or none §	All or none case-series	Absolute SpPins and SnNouts††	All or none case-series	Absolute better-value or worse-value analyses ††††
2a	SR (with homogeneity*) of cohort studies	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups	SR (with homogeneity*) of Level >2 diagnostic studies	SR (with homogeneity*) of 2b and better studies	SR (with homogeneity*) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split-sample § § § only	Exploratory** cohort study with good††† reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity*) of case-control studies		SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible
4	Case-series (and poor quality cohort and case-control studies § §)	Case-series (and poor quality prognostic cohort studies***)	Case-control study, or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

*	By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category.)
†	Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category.)
‡	See note #2 for advice on how to understand, rate and use trials or other studies with wide confidence intervals.
§	Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became
§§	By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in
§§§	Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and
††	An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a
†††	Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits.
††††	Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor reference standards are
†††††	Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more
**	Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the
***	By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or
****	Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (eg 1-6 months acute, 1 - 5

資料2 Reid MC et.al. Use of methodological standards in diagnostic test research: getting better but still not good. JAMA 1995 274:645-51.

Reporting standard	Background	Criteria	Percent meeting standard
Spectrum composition	The sensitivity and specificity of a test depend on the characteristics of the population studied. Change the population and you change these indices. Since most diagnostic tests are evaluated on populations with more severe disease, the reported values for sensitivity and specificity may not be applicable to other populations with less severe disease in which the test will be used.	For this standard to be met the report had to contain information on any three of these four criteria: age distribution, sex distribution, summary of presenting clinical symptoms and/or disease stage, and eligibility criteria for study subjects.	27
Pertinent subgroups	Sensitivity and specificity may represent average values for a population. Unless the condition for which a test is to be used is narrowly defined, then the indices may vary in different medical sub groups. For successful use of the test, separate indices of accuracy are needed for pertinent individual sub groups within the spectrum of tested patients.	This standard is met when results for indices of accuracy were reported for any pertinent demographic or clinical sub group (for example symptomatic versus asymptomatic patients).	9
Avoidance of workup bias	This form of bias can occur when patients with positive or negative diagnostic test results are preferentially referred to receive verification of diagnosis by the gold standard procedure.	For this standard to be met in cohort studies, all subjects had to be assigned to receive both the diagnostic test and the gold standard verification either by direct procedure or by clinical follow up. In case-control studies credit depended on whether the diagnostic test preceded or followed the gold standard procedure. If it preceded, credit was given if disease verification was obtained for a consecutive series of study subjects regardless of their diagnostic test result. If the diagnostic test followed, credit was given if test results were stratified according to the clinical factors which evoked the gold standard	51
Avoidance of review bias	This form of bias can be introduced if the diagnostic test or the gold standard is appraised without precautions to achieve objectivity in their sequential interpretation - like blinding in clinical trials of a treatment. It can be avoided if the test and gold standard are interpreted separately by persons unaware of the results of the other.	For this standard to be met in either prospective cohort studies or case-control studies, a statement was required regarding the independent evaluation of the two tests.	43
Precision of results for test accuracy	The reliability of sensitivity and specificity depends on how many patients have been evaluated. Like many other measures, the point estimate should have confidence intervals around it, which are easily calculated.	For this standard to be met, confidence intervals, or standard errors must be quoted, regardless of magnitude.	12
Presentation of indeterminate test results	Not all tests come out with a black or white, yes/no, answer. Sometimes they are equivocal, or indeterminate. The frequency of indeterminate results will limit a test's applicability, or make it cost more because further diagnostic procedures are needed. The frequency of indeterminate results and how they are used in calculations of test performance represent critically important information about the test's clinical effectiveness.	For this standard to be met a study had to report all of the appropriate positive, negative or indeterminate results generated during the evaluation and whether indeterminate results had been included or excluded when indices of accuracy were calculated.	26
Test reproducibility	Tests may not always give the same result - for a whole variety of reasons of test variability or observer interpretation. The reasons for this, and its extent, should be investigated.	For this standard to be met in tests requiring observer interpretation, at least some of the tests should have been evaluated for a summary measure of observer variability. For tests without observer interpretation, credit was given for a summary measure of instrument variability.	26

資料 3.

スクリーニングと診断用検査

(Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests

<http://www.cochrane.org/cochrane/sadt.htm> より)

Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests:

Recommended Methods, updated 6 June 1996. Available at

<http://som.flinders.edu.au/fusa/cochrane/>

1. 診断検査の系統的再評価に用いる一次研究に求められる情報のチェックリスト

(注釈：チェックリスト内で同じコードの項目は、下記の項目に該当する。)

A. 研究の妥当性における基準：全ての一次研究

- ・ a.検査は適切な標準的診断確定法(reference standard)と比較されているか？
用いられ標準的診断確定法の種類によって研究をカテゴリー分けする
- ・ B.検査は標準的診断確定の方法とは互いに独立に(盲目的に)なされているか？
このカテゴリーは以下のものからなる。
検査も標準的診断確定法も互いに独立になされていた。(最も適切)
検査は標準的診断確定法とは独立になされていたが、逆はそうではない
標準的診断確定法は検査と独立になされていたが、逆はそうではない。
検査も標準的診断確定法も互いに独立にはなされていなかった。(もっとも不適切)
- ・ C.標準的診断確定法を行う患者の選択は検査結果に独立してなされていたか？ (参照バイアスを防ぐ)
- ・ D.検査は他の全ての臨床情報に独立してなされていたか？
- ・ E.標準的診断確定法は、検査の結果でいかなる治療的介入もなされる前になされていたか？ (治療パラドックスを防ぐ)

B. 検査を比較している研究における適切性の追加基準

f.検査の比較は適切なデザインでなされていたか？

このカテゴリーは、以下のものを含む

- 各々の人に全ての検査が独立になされていた(例えば、互いの検査結果には盲目的に) (もっとも適切)
- 無作為に割り振られた人に異なる検査がなされた。
- 各々の人に全ての検査がなされたが、検査は独立的に評価はされていない
- それぞれの人に異なる検査がなされ、その割り振りは無作為ではない。

C. 結果が適用可能かについての基準

C1.臨床的な問題

- ・ a.疾患のスペクトラム(例えば、標準的診断確定法によって"疾患あり"とされた患者の癌の病期分布)
- ・ b.非疾患のスペクトラム(例えば、標準的診断確定法によって"疾患なし"とされた患者

のケースミックス)

- ・ c.セッティング(プライマリ・ケア、三次医療機関、外来、入院など)
- ・ d.検査までの疾患の期間
- ・ e.以前の検査/紹介による選択、例えば、どんな臨床情報を持つ人に検査(それ以前になされた検査を含む)がなされるのか?
- ・ f.併存する疾患
- ・ g.人口動態的情報(年齢、性別、上記の項目と同じように関わる他の因子)

C2. 検査

- ・ h.検査がいかなされたかといった分類。例えば、生化学による検査法の種類
 - ・ i.明確な閾値(カットオフ値)の言及
 - ・ j.検査が容易に行えないあるいは結果が判定できなかったために除外された例の率(%)
 - ・ k.検査の再現性
- D. 研究の質や適用性についての間接的な視点
- ・ a.研究のなされた(あるいは出版された)年
 - ・ b.疾患の有病率
 - ・ c.対象数
 - ・ d.研究デザインが前向きか後ろ向きか
 - ・ e.論文として出版されたか、抄録のみか

チェックリストの注釈

A. a

チェックリストでは“検査(test)”という用語は、他に定義されていなければ、その精確度が評価されている検査を意味し、以前になされた検査、あるいは、標準的診断確定に用いられる検査とは対照のものである。

標準的診断確定(reference standard)は、現在、疾患が存在する、あるいは、ある定義されたフォローアップ期間内に臨床的疾患が判明することによる場合もある。

標準的診断確定法は“完全に適切”であると言ったことは滅多にない。理想的には、異なる研究で用いられた標準的診断確定法はその妥当性(精確度)からランク付けされるべきである。

A. b(測定バイアス)

真ん中二つのカテゴリーのいずれかがもっとも妥当かは検査と標準的診断確定法による

A. c(選択バイアス)

このクライテリア(基準)では、C1.C-f のような興味の対象になっている臨床的問題がみられる対象集団についての明確な定義があることが求められる。興味の対象になっている臨床的問題を有する対象集団についての明確な定義があるかないかは記載されてなければならない。

定義された対象集団の中では、以下のサンプリング戦略(方法)のいずれかによって参照バイアスを防ぐことができる。

- ・ 一連の患者について標準的診断確定法を行う
- ・ 一連の患者の中から無作為に抽出して標準的診断確定法を行う
- ・ 検査が陽性あるいは陰性であった人の中から無作為に抽出して標準的診断確定法を行い、抽出率が異なる場合にはその率で補正する
- ・ 標準的診断確定法で疾患がある、あるいは無いとされた集団について無作為に抽出し検査を行う。

これらのより詳しい情報が欠落している場合には、検査を受けた後に、標準的診断確定法に進まなかった全体の%を記録しておく。

B.e.

中間の二つのカテゴリーにおいて、どちらがもっとも妥当かは、その検査と標準的診断確定法による。

C.2j

追加の有用な情報は対象のなかで除外された人の標準的診断確定法の結果(所見)である。

D.c

サンプルサイズは全体のサンプルサイズではなく、疾患を有する、あるいは有しないグループの最小の方とすべきかもしれない。

参考文献：

1. Arrol, B.A., Schechter, M.T., and Sheps, S.B. The assessment of diagnostic tests: a comparison of medical literature in 1982 and 1985. J Gen Intern Med 3:443-447, 1995
2. Reid M.C. Lachs M.S., Feinstein A.R. Use of methodological standards in diagnostic test research JAMA 274: 6945-651, 1995
3. Lijmer, J.G., Mol B.W. Heisterkamp S. et al. Empirical evidence of design-related bias in studies of diagnostic tests JAMA 282:1061-1066, 1999
4. Bruns D.E., Huth E.J. Magid E. Young D.S. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests Clinical Chemistry 46:893-895
5. Goldman L, Wasson J, Sox H, Neff R. Clinical prediction rules: applications and methodological standards. N Engl J Med 1985;793-799.
6. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature III. How to use an article about a diagnostic test. JAMA 1994;271:389-391.
7. Mulrow CD, Linn WD, Gaul MK. Assessing Quality of a Diagnostic Test Evaluation. Journal of General Internal Medicine 1989;4:288-295.
8. Sheps SB, Schechter MT. The Assessment of Diagnostic Tests: A Survey of Current Medical Research. JAMA 1984;252:2418-2422.
9. Simon, R. and Altman, D.G. Statistical aspects of prognostic factor studies in oncology. Br.J.Cancer 69:979-985, 1994.

資料 4

診断特性に関する論文のチェックリスト

No.	項目	選択肢
	0. 一般項目	
1	セッティング	一般診療所/一般病院/紹介病院(中核病院 or 大学病院)/地域/ボランティア
2	施設間協力	単施設/多施設(<5施設)/多施設(≥5施設)
3	経済的支援	なし/政府系/財団/関連会社/非関連会社
4	検査間比較	単一/複数(2種以上)
5	精確度評価(Accuracy: End points)	感度&特異度/尤度比orオッズ比/ROC/ROC&曲線下面積
	1. 研究実施デザイン	
1	研究実施タイプ	横断的/縦断-後ろ向き/縦断-前向き/RCT/不明
2	検査は適切な標準的診断確定法(reference standard)と比較されているか?	適切/まずまず/不適切
3	検査は標準的診断確定の方法とは互いに独立に(盲目的)になされているか?	互いにマスク/標準的検査にはマスク/目的検査にはマスク/マスクなし
4	標準的診断確定法を行う患者の選択は検査結果に独立してなされていたか?	はい/いいえ/不明
5	検査は他の全ての臨床情報に独立してなされていた	はい/いいえ/不明
6	標準的診断確定法は、検査の結果でいかなる治療的介入もなされる前になされていたか?	はい/いいえ/不明
7	対象集団(患者)の選択(inclusion & exclusion)基準は?	基準が明確/基準が明確とは言えない/記載なし・不明
8	対象集団に対する患者選択については?	
8A	Cohort研究	ある基準(症状・徴候)で選択された患者全て/ある基準の集団から無作為に抽出した患者/不明
8B	Case-Control研究	対照のマッチングが妥当/マッチングが不適切/不明
9	複数検査(index test)の比較は適切なデザインでなされていたか?	全ての患者に全検査を割り付け/患者ごとに無作為に検査の割り付け/異なる患者に異なる検査(非無作為)を割り付け
	2. 検査手法	
1	検査方法についての記載が適切になされているか?	適切/不十分だが再現可能/不適切or記載なし
2	検査の標準物質(reference material)についての記載がなされているか?	適切/あるが適切と言えない/記載なし
3	試料の保存についての記載がなされているか?	適切/あるが適切と言えない/記載なし
4	検査の基準値・カットオフ値・境界値の扱いについての記載がなされているか?	適切/あるが適切と言えない/記載なし
5	検査の再現性(日内・日差変動)についての記載がなされているか?	適切/あるが適切と言えない/記載なし
	3. 結果の記載と統計	
1	対象患者の特性(年齢・性別・受診医療機関・疾患の重症度・併存疾患など)について記載されているか?	適切/まずまず/不適切
2	疾患の有病率について記載されているか?	はい/いいえ
3	確定診断に進んだ例・ドロップアウト例についての記載がされているか?	全例受けた/受けなかった例の詳細が記載/受けなかった例の詳細不明/記載なし
4	結果の主なる表現とその記載は?	
4-1.	2×2表で記載されているか?	はい(複数のカットオフ値で)/はい(単一のカットオフ値で)/いい
4-2(A)	ROC曲線、または、複数のカットオフ値についての尤度比、あるいはオッズ比が記載されているか?	いくつかのカットオフ値による尤度比/ROC&曲線下面積/ROC(曲線下面積なし)/単一カットオフ値の感度・特異度
4-2(B)	ROC曲線下面積、尤度比・オッズ比についてはその信頼区間が記載されているか?	信頼区間が記載/値のみ/記載なし
5	サブグループについての解析がなされているか	はい/いいえ

備考

1-2. 適切：臨床的に考えて最も真に近い、まずまず：最適な確定診断は他にあるが、臨床的に了解できる範囲である。

注：標準的診断確定法は"完全に適切"であることは滅多にない。理想的には、異なる研究で用いられた標準的診断確定法はその妥当性(精確度)からランク付けされるべきである。1-3. "review bias"を防ぐ。結果の判定が、他の検査の結果に独立して(その結果をマスクして)なされているか？ index test が先にされる前向きコホートスタディでは、reference standard がその検査と独立に評価される。reference standard が先行する case-control では、index test がその reference standard の結果と独立して評価されていること

1-4. "work up or verification bias"を防ぐ。 Cohort study では、両方の検査が行われる。あるいは、reference standard が侵襲が強い、あるいは高価の場合には、綿密な観察で。Case control では、index test が先行する場合に、その結果によらずに reference standard がなされていること。Reference standard が先行する場合には、その reference standard を行うことになった臨床的な因子で、index test の結果が層別化がなされている。

1-6. "treatment paradox"を防ぐ。検査結果で治療が行われた後に、reference standard が施行されると、結果に影響する。

1-7. 対象となった subject の組み入れ基準と除外基準が明確に述べられているか

1-8. 組み入れ基準を満たす subject の中での実際に検査がなされた subject の選択についての記載。定義された対象集団の中では、次のサンプリング戦略(方法)のいずれかによって参照バイアスを防ぐことができる。

- ・一連の患者について標準的診断確定法を行う
- ・一連の患者の中から無作為に抽出して標準的診断確定法を行う
- ・検査が陽性あるいは陰性であった人の中から無作為に抽出して標準的診断確定法を行い、抽出率が異なる場合にはその率で補正する
- ・標準的診断確定法で疾患がある、あるいは無いとされた集団について無作為に抽出し検査を行う。

1-8B. 確定診断から選択された患者をケースとする。

2-1. 測定に関する基本的な performance が記載されているか

2-2. 標準物質(一次、二次など)の記載がされているか？

2-3. 採取～測定に至るまでの記載(保存する場合の方法など)がされているか？

2-4. indeterminate(灰色) test results の取り扱いを含む

2-5. 検査の再現性の検証 (測定方法の reproducibility)

3-1. Spectrum composition 他施設でも結果が活用できるか？(外挿性が OK か？)

3-2. 算出可能であれば良い

3-3. 対象 subject で最終的に結果に含まれなかった例数とその理由について記載され

ているか？

3-4. precision of results for test accuracy : 分布図で示される場合には、その実数が読みとれる場合には同様に扱う

2 × 2 表で表示されているもの(精確度・信頼区間については計算可能)と ROC 曲線、尤度比については同格と扱う。実際には ROC 曲線下面積がその信頼区間とともに表示されている場合が良い場合(複数検査の比較時)と、複数(あるいは単数)カットオフ値での尤度比が記載されている場合(実際の臨床に活用する)とで異なる。

3-4-1 と 3-4-2 の両方が当てはまる場合には、どちらか一方を選択する。

3-5. 適切に層別化された集団での test accuracy が記載されている？

資料 5 Summary ROC 曲線についての主な論文

1 Shapiro, D.

Issues in combining independent estimates of the sensitivity and specificity of a diagnostic test. Acad Radiol 2:S37-S47, 1995.

診断テストについての複数の研究を統合する際に、重みづけ最小二乗法(Moses & Littenberg)、最尤法(Kardauns)や潜在スケールロジスティック回帰モデル(Rutter & Gatsonis)などの現行で活用可能な統計的手法についての秀逸なレビュー論文である。また、ROC における各研究のポイントにおける信頼域についての提案している。

2. Kardaun, J.W.P.F. and Kardaun, O.J.W.F.

Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. Meth.Inform.Med. 29:12-22, 1990.

感度・特異度への logit 変換を行った先駆的な論文で一定のオッズ比をとることを前提に閾値の変化を許容している。それぞれの研究を信頼域分布を有するロジスティック ROC 領域に上手く展開している。

3-1. Moses L.E., Littenberg B, Shapiro D.E.

Combining independent studies of a diagnostic tests into a summary ROC curve: data-analytical approaches and some additional considerations. Stat Med 12:1293-1316, 1993

3-2 Littenberg, B. and Moses, L.E.

Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 13:313-321, 1993.

log オッズ比 (=logit(TPR) - logit(FPR))、すなわち、識別能がその陽性基準で変わりうるかをみるため、logit(TPR) - logit(FPR)と logit(TPR) + logit(FPR)との間の解析を用いている。もし、陽性基準で変わらなければ、その結果として log オッズ比は、まとめの指標になる。かれらは、非重み付け、重み付け、そしてロバスト版を考慮している。

4. McClish, D.K.

Combining and comparing area estimates across studies or strata. Med Decis Making 12:274-279, 1992.

全体の ROC 領域の面積の推定値を得るために(分散の逆数で)重み付けされた和を用いている。これはオッズ比を結合しまとめた Mantel-Haenszel による推定の変法と考えることができる。

5-1. Tosteson AN, Begg CB.

A General Regression Methodology for ROC Curve Estimation. Medical Decision Making 8:204-215,1988

ROC 曲線に Fit するために ordinal regression models を用いるもの

5-2. Tosteson AN, Weinstein MC, Wittenberg J et al.

ROC Curve Regression Analysis: The Use of Ordinal Regression Models for Diagnostic Test Assessment. Environmental Health Perspectives 102(suppl 8) :73-78, 1994

共変量(covariate)と Regression model と ROC 曲線下面積についての関連を述べたもの

6. Rutter, C.M. and Gatsonis, C.A.

Regression methods for meta-analysis of diagnostic test data. Acad Radiol 2:S48-S56, 1995.

Fixed や random モデルによる潜在スケール(latent scale)を用いたロジスティック回帰法を提示している。

資料 6

[1] CASP ワークショップ参加報告

はじめに

CASP は EBM の啓蒙を目的として、医師だけでなくコメディカルや一般市民を対象とし、文献の検索や評価のノウハウを実践的に学ぶ、参加型ワークショップを中心とする英国で始まった活動で、日本でも CASP Japan により既に 10 回程度開催されています。今回は臨床 EBM 研究会第 5 回公開セミナーの一部として実施されたので、CASP ワークショップ以外の内容も含めまとめておきます。

特別講演 「診療ガイドラインはどのように作られるのか」

Dr. ロジャー・ウィン (U.T.M.D. アンダーソン癌センター)

全米の癌治療センターをネットワークする NCCN (National Comprehensive Cancer Network) が整備を進めている、総合的な癌治療ガイドラインの作成手順の概説。エビデンスの得られない場合のコンセンサスの作り方の他、ガイドラインの普及率や導入効果をモニターしながらきめ細かく改訂している点や、フローチャート形式にしたり医師用とは別に患者説明用も揃えるなど、利用者の便宜を強く意識している点が注目された。厚生省の医系官僚から、これだけ充実したガイドライン整備の原動力は何かとの質問があり、米国では医療保険の民営化に対抗するため医師側に理論武装の必要が生じたとの指摘があった。

教育講演 1 「今日からはじめる EBM : EBM 実践のためのノウハウ集」

名郷直樹先生 (作手村国民健康保険診療所 所長)

講師は日常診療でコクラン等のデータベース利用を実践していることで有名な医師で、豊富な経験に基づく実用的なアドバイスがあった。講師の底抜けに明るく前向きなキャラクターとともに、診療を EBM に適合させるのではなく、あくまで患者とのコミュニケーションの壁を取り除く道具の一つとして、EBM の限界を熟知しつつ役立つ部分だけをうまく利用するという姿勢が印象に残った。

教育講演 2 「CASP の実際 : 楽しい批判的吟味」

福岡敏雄先生 (名古屋大学医学部附属病院 救急部・集中治療部)

CASP の概説であったが、できるだけ身近な例や平易な言葉を多用するなど、専門家以外への参加者の広がりを目指す姿勢が強く感じられた。また、日本において CASP が短期間でこれだけの規模に育った最大の要因は、講師個人の気さくで明るいキャラクターの魅力であることが伺えた。

スモールグループチュートリアル

(CASP 以外に治療 / 副作用 / 判断分析 / ガイドラインのグループに分かれ同時進行)

CASP ワークショップは全体の受講者 77 名のうち 34 名が選択するという盛況ぶりであった。ただし医師は少数派で、むしろ看護婦、薬剤師、学生、医療商業誌編集者あるいは患者団体世話人の方々が積極的に参加している姿が印象的であった。また全体の進行は CASP Japan を率いる福岡 敏雄先生が務められたものの、各 10 人程度のグループディ

スカッションでチューターを務めていたのは医師以外の方々であった。今回の内容は、帝王切開に対する抗生剤の予防投与につき、コクランの該当システムティックレビューを対象に、CASPのチェックシートに沿って、想定された事例にその結論を適応するのが妥当かどうかを検討するというものであった。GLMと比較し求められるプロダクトのノルマは緩やかで、使用する論文や豊富な資料がすべて翻訳されているなど、予備知識の少ない受講者にもできるだけ敷居が低くなるように配慮されており、膨大な労力と長期に渡る細かなノウハウの積み重ねが感じられた。実際にはあまりにも受講者の背景が多様なため、議論のレベルには限界があったが、チューターの巧みな誘導により脱落者は殆どなく、中には同じ課題につき全く異なる見方が聞けることを高く評価する声もあった。

運営について

費用的には受講費（一人1万円、食費宿泊費別）で大体賄い、足りない分を臨床EBM研究会が数社から支援を受けているようである。ただし、CASPのチューター等については、実費もしくは一部持ち出しといったボランティアベースとのことであった。CASPの運営については、現在CASPの事務負担の殆どを福岡敏雄先生個人が担っており、人的ネットワークの拡充による作業負担の分散が次の課題とのことであった。

まとめ

EBDセミナー（仮称）をCASP形式で実施するためには、費用の制約はそう大きくないが、受講者を満足させられるだけの内容やそれを反映した十分な教材と共に、相応の適性、能力および意欲を備えたチューターを揃える必要がある。EBM関連の活動に強い関心を抱いているのは、クリティカルパスの作成を迫られている看護婦、臨床薬剤師への転換を迫られている薬剤師、医療不信を募らせている患者などであり、医師特にタコ壺的専門医の関心は低いので、セミナーの対象を学会員に限った場合には、何らかの工夫をしないと受講者が殆ど集まらない恐れがある。

以上

(http://ebd.umin.ac.jp/new/casp_r.html)

[2] 国際学術集会出席報告書（第6回アジア臨床病理学会）

=====

2000年10月28日

文部省 EBD 研究班員 各位

研究分担者 西堀眞弘

国際学術集会出席報告書

集会名称 第6回アジア臨床病理学会

開催期日 2000年10月11日（木）～同13日（土）

開催場所 韓国 プサン市 Haeundae（海雲台）Grand Hotel

(1) 本学術集会の概要

アジア各国の臨床病理学の発展を促進するために2年に一度開催される国際学術集会で、今年の世界14カ国より約360名が参加し、展示は35社と盛大であった。

(2) 演題発表について

日本における臨床検査医学領域のインターネット利用の概要を「IT Strategies of Laboratory Medicine in Japan」という演題で口演発表(後掲資料参照)した。主催者側の不手際により、座長が不在というハプニングのため発表時間が限られ、質疑応答が省略されてしまったが、後程韓国のCho Han Ik教授より、今後特にEBLM関連の活動について協力関係を持ちたい旨の申し出があった。

以上

・発表内容

(<http://mn.umin.ac.jp/work20001013.html>)

=====

[3] 国際学術集会出席報告書(The 8th. International Cochrane Colloquium)

(http://ebd.umin.ac.jp/kaigi/cochrane_rep.html)

国際学術集会出席報告書(The 8th. International Cochrane Colloquium)

2000年10月29日

研究代表者 石田 博

研究分担者 西堀真弘

開催期日:2000年10月25日(火)~同29日(日)

開催場所:南アフリカ共和国 ケープタウン市 Good Hope Centre

議事録

(1) 本学術集会の概要

年に1度開催されるCochrane Collaboration最大の学術集会で、今回は8つのPlenary Session、55の口演発表、155のポスター発表の他、40のさまざまなトピックのワークショップに加え、開会前の2日に渡り様々な関連セミナーが組まれていた。会員制ではなくボランティアベースのゆるやかな組織ではあるが、300名を越すと思われる参加者はそれぞれ高い問題意識を持って集まっているようで、多くは会期中会場を離れず熱心に参加していた。

出席者は発祥の地英国をはじめとする西欧・北欧諸国、米国、カナダ、オーストラリア、中国、韓国等に加え、開催地を反映してアフリカ諸国およびフィリピンなどの途上国からも幅広く参加していた。日本からの出席者は我々を含め4~5名と見られ、南米、東南アジア、東欧諸国からは少なかった。

テーマとしてはメタアナリシスあるいは EBM 実践の経験やノウハウの他、不均一なデータ (heterogeneity) の扱い方、無作為抽出試験の難しい場合 (non-randomized studies) の扱い方、統計値でなく個々の元データ (individual patient data; IPD) に基づくメタアナリシス、前向き試験に基づくメタアナリシス (prospective meta-analysis; PMA)、患者 (consumer) への配慮や患者の直接参画などが新たなテーマとして注目されていた。

なお、出席できたセッションでは特に資金面での質疑応答が目につき、補助金を得た事業は部分的に盛んになるが、それが途絶えるとじり貧になるという、ボランティアベースによる運営に共通する問題点が、設立後 7 年目を迎え早くも表面化しつつあるように感じられた。

(2) 演題発表について

ICPC/C-SRLM および文部省 EBD 研究班の概要を「A Systematic Approach to the Evidence Based Laboratory Medicine in Japan」という演題でポスター発表 (後掲資料参照) した。参加者から特に質問はなかったが、日本の総力を結集した試みに期待する声寄せられた。

(3) IFCC/C-EBLM メンバーとの話し合いについて

IFCC/C-EBLM の一員である Andrea Rita Horvath 女史と、10月27日(木)13時より約1時間にわたり会談した。Rita 氏からは IFCC 会長の意向および C-EBLM の活動状況等の説明があり、当方からは当研究班および ICPC/C-SRLM の活動状況を、発表ポスター、各々のホームページおよび SRLM 用チェックリスト試案を提示しつつ説明した。

Rita 氏は当研究班の進めているチェックリストの作成と EBLM データベースの構築について特に関心を示し、彼女の目的とする Cochrane Collaboration と IFCC の協力関係の模索において、重要な要素として考慮に入れたいとの意思表示があった。

(4) Cochrane Collaboration の business meeting への出席について

10月28日(金)12時30分より会場内の会議室で開催された The Cochrane Methods Group on Screening and Diagnostic Tests の business meeting に Rita 氏と共に出席した。出席者は我々を含めて12人で、以下の議題が話し合われた。

1. SRs of diagnostic accuracy in Cochrane
2. Other databases (Medion, DARE, IFCC)
3. software development
4. STARD
5. How do we handle queries from other entities
6. Links to other bodies e.g. IFCC
7. Updating the website

この中の 6.において Rita 氏より日本の状況が説明され、質疑応答の後、当研究班とし

てこの Methods Group が進めようとしている EBLM データベース構築への協力を表明した。その他の議事の具体的内容は省略するが、後日事務局より参加者のメールアドレスのリストと共に、議事要旨が配布される予定なので、その際に改めて報告する。

なお、SRLM のガイドラインの新バージョンが提示されるとの情報があつたが、実際には提示されなかつた。むしろチェックリストの作成、EBLM データベースの構築などと共に、未だに考え方の整理をしている段階にあると思われた。また実際にレビューを行う Collaborative Review Group が臨床検査医学分野の論文を対象として活動するためには、この Method Group においてガイドライン等が固まるのを待つ必要があるとの感触であつた。

(5) その他

開催地ケープタウンは大変過ごしやすい気候で、海の幸などの食材にも恵まれているうえ、テーブルマウンテンや喜望峰などの見所も多いが、会議開催地や観光地としての開発は今一步の感である。また会議場やホテルには警備員が配置され、学会主催者からも車の移動が勧告されるなど、アパルトヘイトの廃止によって流動化した人口が都市に流入するにつれ、治安は悪化しつつあるとの印象を受けた。

以上

[参考資料] ポスター発表の内容

The 8th International Cochrane Colloquium

A Systematic Approach to the Evidence Based Laboratory Medicine in Japan

2000.10.25-29, Cape Town, South Africa

Masahiro Nishibori (Tokyo Medical and Dental University, Japan), Tadashi Kawai, Takashi Kanno, Masayuki Kambe, Kiyooki Watanabe, Tsuguya Fukui, Haku Ishida

Introduction

- ・ Evidence-based laboratory medicine (EBLM) requires a more specialized approach than usual evidence-based medicine, because;

- # Unlike the effectiveness of therapeutic measures which can be evaluated by comparing only two groups, there will be almost infinite combinations of various diagnostic tests and various diagnostic phases in the clinical course.

- # Unlike the medication which has direct effects to the patient outcome, many steps should be taken to analyze effects of performing laboratory tests to the final outcome of the patient.

- # There are many factors which cause various statistical biases of the laboratory data.

Purpose

- To organize a comprehensive research group of the leading members of all fields related to the EBLM.
- # From the experts group of laboratory informatics
- # From the committee of guidelines for utilizing laboratory tests
- # From the committee for standardization of laboratory tests
- # From the experts group of statistics
- # From major academic societies in the field of laboratory medicine
- To establish appropriate guidelines proper to the EBLM for inspection of usefulness of laboratory tests.
- To promote continuous application of EBLM in practice.

Methods

- The Committee on Systematic Reviewing in Laboratory Medicine (C-SRLM , <http://ebd.umin.ac.jp/srlm/>) was organized by International Clinical Pathology Center (ICPC) as a non profitable expert group.
- It consists of the following four teams
- # Methodology Team
- # Selection Team
- # Statistics Team
- # Management Team
- Also it collaborates with the Evidence Based Diagnosis (EBD) Research Project Team (<http://ebd.umin.ac.jp/>), which was organized by Japanese government.

C-SRLM in detail

- Chair
- # T. Kawai (Director of International Clinical Pathology Center; ICPC)
- Vice Chair
- # T. Kanno (President of Japan Society of Clinical Chemistry; JSCC)
- Methodology Team
- # Chief: M. Kambe, chairman of the Committee of Laboratory Informatics, the Japan Society of Clinical Pathology; JSCP
- # It prepares the standard methodology manual for systematic reviewing (meta-analysis) and the requirements for publications which deal with clinical usefulness of diagnostic tests.
- Selection Team
- # Chief: K. Watanabe, chairman of the Committee of Utilization of Laboratory Tests, the Japan Society of Clinical Pathology; JSCP

It selects the subjects suitable to be studied in this project.

- Statistics Team

Chief: T. Fukui, Professor of Clinical Epidemiology, Kyoto University, one of the leading researcher in the field of the evidence based medicine; EBM

It studies and proposes the most suitable statistical methods to be applied in Systematic Reviewing in Laboratory Medicine (SRLM).

- Management Team

Chair and Vice Chair of C-SRLM

Leaders of Methodology, Selection and Statistics Team of C-SRLM

K. Kawano (President, Japanese Association of Clinical Laboratory Physicians; JACLaP & Japanese Committee for Clinical Laboratory Standards;JCCLS)

I. Sakurabayashi (President, the Japan Society of Clinical Pathology)

K. Nakahara (Professor, Department of Laboratory Medicine, Tokyo University)

It manages the C-SRLM activity and communicates with national and international organizations.

EBD Research Project Team

- Dedicated to studying 'How to Establish Systematic Review of Diagnostic Tests Aimed at Efficient Medical Practice of Good Quality'

- Supported by Grant-in-Aid of Year 2000 from the Ministry of Science, Education, Sports and Culture of Japan (Grant No.12897028)

- Chief researcher : Haku Ishida, Kawasaki Medical College, Japan

- It provides assistance of highly-advanced technologies to the C-SRLM in preparing:

a mailing list for efficient discussions.

a bilingual WWW site to accumulate the database collected through this project and finally to make them easily accessible to the public.

Results at Present

- The initial EBLM oriented research have been successfully conducted.

The First Study Group on SRLM on High Sensitive CRP Measurement dedicated to the multi-centers clinical study on Japanese populations in the cases of neonatal infections and cardiovascular and cerebrovascular diseases

- Its collaboration with the international organizations is ongoing.

The Cochrane Methods Group on Screening and Diagnostic Tests

The Committee of Evidence-based Laboratory Medicine of the International Federation of Clinical Chemistry (IFCC/C-EBLM)

- 発表内容

(<http://mn.umin.ac.jp/work20001026.html>)

・ Minutes of the Meeting of Methods Group on Meta-analyses of diagnostic and screening tests. Capetown, 27/10/2000

(http://ebd.umin.ac.jp/kaigi/Minutes_MG_diag_capetown.html)

[4] 国際学術集会出席報告書(The 2nd China-Japan-Korea Joint Symposium on Medical Informatics)

(http://ebd.umin.ac.jp/kaigi/JKC_Mediinfo.html)

国際学術集会出席報告書(The 2nd China-Japan-Korea Joint Symposium on Medical Informatics)

2000 年 12 月 4 日

研究分担者 西堀真弘

開催期日: 2000 年 12 月 2 日 (土)

開催場所: 大韓民国 済州 Hyatt Regency Cheju

(1) 本学術集会の概要

医療情報学分野ではかなり以前から日中間の交流が続けられており、そこへ韓国が合流して昨年より新たに始まったシンポジウムである。第 1 回は中国の蘇州で開催され、来年は医療情報学会の期間中に東京で開催される予定である。

今回は 1 日の枠中で 66 の演題すべてが 3 つのトラックに分かれて口演発表され、その他に 3 題の Plenary Lecture があった。参加者は 100 余名で、韓国が最も多く、次いで中国、日本は最も少なく 10 名程度であった。また昨年医療情報学会が設立されたモンゴルと、オーストラリアからゲストが来ていた。

演題の内容は電子カルテ、標準化、セキュリティ、画像処理、病院情報システム、遠隔医療、看護情報、薬剤情報、教育、生物工学など、日本とは大差なく、日本と同様に看護および薬剤領域からの熱心な参加者が目立った。発表のレベルは必ずしも高くないが、昨年に比べて全体的に質の向上が著しく、特に韓国からの演題は日本より進んでいるものも見られた。

(2) 演題発表について

ICPC/C-SRLM および文部省 EBD 研究班の概要を「A Systematic Approach to the Evidence Based Laboratory Medicine」という演題で口演発表した。フロアからは高感度 CRP の標準化と、patient outcome 改善の効果の実績について質問があり、また韓国で始めて医療情報学講座を開設した Kyungpook National University の Yun Sik Kwak 教授より、本プロジェクトを高く評価し、積極的に関与したい旨のコメントがあった。また終了後には同講座の Hune Cho 助教授と Gachon 医科大学の Byeong-Gook Lee 助教授より、後でコンタクトしたい旨の申し出があった。またソウル大学校医科大の Han-Ik Cho 教授

からは、来年の韓国医学会で提言を予定している EBM への取り組みについて、コメントを求められた。

(3) 各国の EBM の現状について

今回 EBM 関連の演題は他になかったが、多くの発表内容中で言及されているため、これらの国々にも認識は十分に浸透し、特に医学教育への利用に関心が向けられているように感じられた。ただし、具体的な対応はまだこれからという印象であった。

・発表内容

(<http://mn.umin.ac.jp/work20001202.html>)

[5] International Symposium on "EBM and Policy Development" 参加報告

=====

2000年2月23日

文部省 EBD 研究班員 各位

研究分担者 西堀真弘

International Symposium on "EBM and Policy Development" 参加報告

開催期日 2001年2月16日(金)

開催場所 都ホテル東京、明日香

主催 平成 12 年度厚生科学研究「日本における EBM のためのデータベース構築及び提供利用に関する調査研究班」(研究代表者：丹後俊郎)

Part I: Meta-analysis (chairperson: Toshiro TANGO)

1. An introduction to meta-analysis (Toshiro TANGO, Director of the Library, NIPH; Div. of Theoretical Epidemiology, Dept. of Epidemiology, NIPH)

本シンポジウム全体の流れとして、Cost-effective な医療のための EBM、その EBM を支える Meta-analysis、とそれぞれを位置づけ、Part I でその解説、Part II で Policy Development への応用を論じたいとの前説の後、30分程度の Meta-analysis の概説があった。その中では、sensitivity analysis (発表年度やサンプル数により層別に再分析して比較する)により、集めたレビュー内に潜在するさまざまなバイアスを検出できるという話が興味深かった。また予備知識のなかった内容(例えば Baysien Model による meta-analysis など)については、項目の羅列のみであったため、あまりよく理解できなかった。

2. How to conduct meta-analysis (Joseph LAU, Co-Director, New England Cochrane Center; Professor and Director, New England Medical Center)

まずメタアナリシスの必要性や有用性が豊富な実例で分かりやすく説明され、演者発案

の cumulative meta-analysis (通常時系列で上から順に論文毎のオッズ比±信頼区間をプロットするところを、各論文を既発表の論文とともにメタアナリシスして得られたオッズ比±信頼区間をプロットしていくと、各論文のオッズ比がばらついていても、早期に収束点が明らかになる)が、急性心筋梗塞に対する血栓溶解療法の有用性を、その領域の専門家よりはるかに早く指摘した実績などが示された。

次いで、Cochrane レビューのメタアナリシスの結果と大規模介入試験の結果が3割以上で一致せず、またこれが大規模介入試験間の不一致率と同程度であることなど、現時点でのメタアナリシスの限界が示された。この原因として演者は、実は患者群はそれぞれ独自の分布を持つ複数のサブグループの集合体であり、サンプリングがサブグループごとに均一あるいはランダムに行われないために、メタアナリシスの結果、特定のサブグループの分布が強調されたり、また各々の分布そのものが歪んだりしている可能性を指摘した。

Part II: EBM and Policy Development (chairperson: Kenji HAYASHI)

3. EBM movement in China: The role played by the Chinese Cochrane Center

(Liu Ming, Deputy Director of Chinese Cochrane Center; Professor of Neurology, West-China University of Medical Sciences, Chengdu)

中国では1999年2月に、四川省の省都である成都市の West-China University にコクランセンターが開設された。演者が受けた英国コクランセンターでの研修を機に、学内の有力者や政府の支持を獲得して開設に至った経緯や、学界だけでなく、医療機関や学術出版界をも巻き込んだ、いわば全方位的な EBM 活動を精力的に進めている現状が紹介され、改めて中華パワーを印象付けた。

4. Policy and politics on EBM in Japan (Kiichiro TSUTANI, Associate Professor, Dept of Clinical Pharmacology, Div. of Information Medicine, Medical Research Institute, Tokyo Medical and Dental University)

最初に、厚生労働省の役人に講演を断られたための代役ではあるが、特に中国における EBM の進展ぶりと、我が国の医師会の無理解に危機感を募らせて引き受けた、とのコメントがあり、いよいよ津谷先生もコクランセンターの開設を本気で考え始めたとの印象を受けた。

本論では、まず日本における EBM の歴史を年表で振り返った後、「EBM に基づく診療ガイドラインは診療行為を制限するもの」および「EBM は万能」という2つの誤解を解消し、臨床医・行政・患者等各々にとって真に役立つ意思決定サポートシステムを確立すべきである、との指摘があった。

また、欠陥のあるエビデンス(栄養バランスの改善による脚気予防効果を実証するため2隻の戦艦の異なる時期の航海を対照と介入事例に選んだ)を根拠に結果的には正しい選択をした(主食を白米から麦飯にした)イギリス学派の高木兼寛と、エビデンスの欠陥を指摘した(比較対照が均一でない)が結果的には誤った選択をした(日清・日露戦争で多

数の兵士を脚気で死亡させた)ドイツ学派の森林太郎(森鷗外)を対比させ、正しい意思決定が一筋縄では行かないことを強調した。

5. Experience of EBM development policy in the United States (Lau Joseph, Co-Director of New England Cochrane Center; Professor and Director, New England Medical Center)

これまで喧伝されてきた DRG/PPS 指向の情報ではなく、米国において行政、医療市場、専門医会、学会、患者団体等各方面で展開されている、本来の EBM 活動が具体的に紹介された。米国においても、医療の効率化を求める圧力が EBM の主な推進力となっているが、初期に作られた脊椎手術に関するガイドラインが手ひどく批判された経緯からか、EBM evidence と診療ガイドラインの違いとして、前者は決して clinical recommendation をしてはいけない、と強調されていたのが印象に残った。

また、望ましいメタアナリシスの実施例として、1万以上の文献をリストアップし、500以上の抄録を読み、100以上の本文を検討し、20以上の論文データを集計することが必要として、1件25万ドル以上の費用見積もりが提示され、個人的努力の限界を感じた。

なお、米国にはメタアナリシスの理論および技術面で個人的に主導している研究者はいるものの、EBM 活動全体を主導する核は見当たらず、また以前は4つあったコクランセンターも運営負担の問題で次々と閉鎖され、近い将来、演者の運営するセンターだけになるであろうとの見通しが示された。

6. Perspective of evidence-based community health guide (Kenji HAYASHI, Director, Department of Demography and Health Statistics, NIPH)

演者は evidence-based community health を EBM と並立する別の枠組みと位置づけ、主に米国における介入試験の実施手順の紹介があった。ただし「後者と異なり前者には RCT の手法が適用できない」などの説明があり、意味がよく理解できなかった。

7. Panel Discussion of Part II

いくつかの追加質問やコメントの後、津谷先生から中国の先生に「コクランセンターを自施設で立ち上げるのに成功した要因は何か」との核心に迫る質問があり、議論が白熱した。結論としては、やはりその成否は高い意欲を持った個人の自発的努力にかかっており資金的裏付けなく組織としての持続的活動を保証することは難しい、という認識で一致した。

参考文献

1. Walraven CV, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA* 1998 280: 550-8.
2. Lundberg GD. The need for an outcomes research agenda for clinical laboratory testing. *JAMA* 1998 280: 565-6.
3. 日本臨床病理学会 日常初期診療における臨床検査の使い方小委員会編 DRG/PPS 対応臨床検査のガイドライン(第二次案) 日本臨床病理学会, 2000
4. 名郷直樹 EBM の必要性と情報環境 中嶋 宏編 EBM のための情報戦略 中外医学社 2000
5. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial *Control Clin Trials* 1981; 2(1):31-49
6. Chrzanowski RS, Paccaud F. Study design for technology assessment: critical issues *Health Policy* 1988;9(3):285-96
7. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12
8. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA* 1995 274:645-51.
9. Lijmer JG Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999 282: 1061-6.
10. Sheps SB, Schechter MT The assessment of diagnostic tests. A survey of current medical research. *JAMA* 1984;252(17):2418-22
11. Mulrow CD, Linn WD, Gaul MK, Pugh JA. Assessing quality of a diagnostic test evaluation. *J Gen Intern Med* 1989;4(4):288-95
12. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994 Apr 15;120(8):667-76
13. Bruns DE, Huth EJ, Magid E, Young DS. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. *Clin Chem* 2000;46(7):893-5
14. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine 2nd. Ed. Churchill Livingstone, 2000
15. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134(8):663-94
16. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the

- quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999;354(9193):1896-900
17. 青木 仕 構造化抄録の基礎知識 中嶋 宏編 EBM のための情報戦略 中外医学社 2000
 18. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *N Engl J Med* 1997; 336(4):309-15
 19. Egger M. et al ed. *Systematic reviews in Health Care* BMJ Book 2000
 20. Cooper H. et al ed. *The Handbook of Research Synthesis* Russel Sage Foundation 1994
 21. Petitti DB *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis* Oxford University Press 2000
 22. Midgette AS, Stukel TA, Littenberg B. A meta-analytic Method for Summarizing Diagnostic Test Performances: Receiver-operating-characteristic-summary Point Estimates *Med Decis Making* 1993;13;253-257
 23. Shapiro, D. Issues in combining independent estimates of the sensitivity and specificity of a diagnostic test. *Acad Radiol* 2:S37-S47, 1995.
 24. Kardaun, J.W.P.F. and Kardaun, O.J.W.F. Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *Meth.Inform.Med.* 29:12-22, 1990.
 25. Vamvakas EC. Meta-analyses of studies of the diagnostic accuracy of laboratory tests: a review of the concepts and methods. *Arch Pathol Lab Med* 1998;122(8):675-86
 26. Moses L.E., Lettenberg B, Shapiro D.E. Combining independent studies of a diagnostic tests into a summary ROC curve: data-analytical approaches and some additional considerations. *Stat Med* 12:1293-1316, 1993
 27. Littenberg, B. and Moses, L.E. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making* 13:313-321, 1993.
 28. Rutter, C.M. and Gatsonis, C.A. Regression methods for meta-analysis of diagnostic test data. *Acad Radiol* 2:S48-S56, 1995.
 29. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;117(1):167-78
 30. Easterbrook PJ, Berlin JA, Copalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991.;337: 867-872
 31. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 6:11-29,1987
 32. Dickersin K, Chan S, Chalmers TC, et al. Publication bias and clinical trials. *Control*

- Clin Trials. 1987;8:343-353.
33. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med*. 1987;6:11-29.
 34. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc, Series A*. 1988;151:419-463.
 35. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst*. 1989;81:107-114.
 36. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA*. 1992;267:374-378.
 37. Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. *JAMA*. 1990;263:1401-1405.
 38. Deeks J. Systematic reviews of evaluations of diagnostic and screening test. In *Systematic Reviews in Health Care*. 2nd. Ed. BMJ Publishing 2001
 39. Hethrington J et al. Retrospective and prospective identification of unpublished controlled trials: Lessons from a survey of obstetricians and pediatricians. *Pediatrics* , 1989 ;84:374-380
 40. McAlister FA, Straus SE, Sackett DL. Why we need large, simple studies of the clinical examination: the problem and a proposed solution. *Lancet* 1999 Nov 13;354(9191):1721-4
 41. Yasue H. et al. Effect of aspirin and trapidil on cardiovascular events after myocardial infarction. Japanese Antiplatelets Myocardial Infarction Study(JAMIS) Investigation. *Am J Cardiol* 1999; 83:1308-1313
 42. Oosterhuis WP. Niessen RW. Bossuyt PM. The science of systematic reviewing studies of diagnostic tests. *Clinical Chemistry & Laboratory Medicine*. 2000; 38(7):577-88,
 43. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104(4):374-80
 44. Price CP Evidence-based laboratory medicine: supporting decision-making *Clin Chem* 2000;46:1041-50
 45. 三宅一徳他 ; 臨床検査 Evidence Database 構築上の問題点と改善の試み 臨床病理,47(補):p285, 1999
 46. 古川壽亮 エビデンス精神医療, p109-146, 医学書院 2000
 47. Peirce JC, Cornell RG: Interating Stratum-specific likelihood ratios with the Analysis of ROC curve. *Med. Dis Making* 1993; 13: 141-151,.
 48. 市原清志ら 疾患別症例データベースからの動的知識生成法と診断支援 臨床病理

1999;47 : 850-859,

49. Guyatt GH, Oxman AD, Ali N, Willan A, Mcilroy W, Patterson C. Laboratory Diagnosis of Iron-deficiency Anemia J Gen Intern Med 1992;7:145-153