

ここで主題から脱線するかもしれないが、重要な点として、どうして、これらのエビデンスがないのかについて、若干考察してみたい。著者は歯科医師のため、歯科疾患を前提としてしまうが、この問題は、外科系全体にあてはまることだと考えている。そもそも、一般内科系の先生にとっては、いわゆるコモンディゼーズ (common disease、一般的な病気) である歯科疾患でエビデンスがないことは信じられないであろう。しかし、実際は、介入の多くが薬剤でなく手術(歯を削るなど)の侵襲的治療であるためにランダム比較試験が行うことが困難である。さらにエビデンスがないことになった理由として、表2のようないろいろな理由がある。

この中で、「症例数が少ない」ということに関しては、そんな稀な疾患に対して、はたして診療ガイドラインが必要かという根本的な問題があるが、これは本稿では取り扱わない。しかし、症例数がある程度多くて、その分野ではメジャーな疾患にも関わらず、ランダム比較試験のように厳密に選択基準や重症度をそろえて臨床試験とすると、施設ごとの症例数が少なくなり、研究の質が下がることで、臨床試験が困難である場合も多い。また、臨床の場面では、症例数が多いにもかかわらず、厳密な選択基準を設けた研究の場面では、症例数が極端に少なくなる場合もある。歯科疾患も同じであり、たとえば、カリエス(虫歯)と言っても、同じ部位で、同じ大きさ・深さで、同じカリエスリスクで、同じ年代とまで対象を限定すると、多いカリエスという疾患に

も関わらず、一つの施設での症例数がほとんどないということになる(それだけ、多様な疾患である)。さらに長期予後で考えると、生活習慣が直接大きく関与するので、生活習慣(砂糖の摂取や歯ブラシの回数)で分類すると個人個人まで分類しなければならなくなってしまう。このように、たとえば胃がんの手術をとっても、プラセボ手術が可能かどうかだけでなく、厳密に体格や術者の腕(手術経験)まで考えると、ランダム比較試験を行うのは困難であることが理解できるはずである(近年は、倫理的観点より、さらに困難となっている)。

実は、このことが診療ガイドラインを作成する時に問題になる。すなわち、臨床で考えられる状況を細かく考えて厳密に対象疾患を定義してしまうと、エビデンスがなくなるということである。しかし、臨床的かつ学問的に許容されないほど、定義をあいまいにしてしまうのも、専門医としては納得がいかない場合に、具体的な解決案が見いだせないことになってしまう。

最後に、コクランレビューの結論によく書かれている、「no evidence」は、ランダム比較試験や準ランダム比較試験がない場合に、観察研究まで調査を広げずに、定型文的に記載されていることが多いので、本稿の定義と異なるし、今後は、観察研究まで調査することが望まれる。少なくとも、学会として、診療ガイドライン作成を数か月遅らせてでも、事例を集めて症例集積研究を行うべきであると考える。

表2: 質の高いエビデンスが作られてこなかった理由

-
- ・研究者の怠慢
 - ・症例数が、もともと少ない
 - ・障害度・重症度などを細かく分類して、その一つ一つで介入が異なることにより症例数が減る

(想定する患者(対象)を厳密に絞りすぎた場合・定義や重症度を一致させ均一な集団とするこ
とによって内部妥当性を高めすぎた場合)

- ・倫理的にランダム比較試験が困難
- ・手技に個人(オペレーター)の差が大きい
- ・環境要因など他の因子(特に生活習慣)の影響が大きい
- ・資金がない(民間企業の参入がない)
- ・古くから行われていた治療でありルーチン化されていた

3. そもそも、「エビデンスがない場合」に推 奨文を作るべきか？

近年、診療ガイドラインは、ただエビデンス
をまとめるだけでなく、できる限りEMBの実践
に役立つ資料を作るため、その作成方法が
議論されている。その代表が2000年に、
GuyattやOxmanが中心となってEBMの方法
論に最適化した診療ガイドライン作成ルール
であり、各学会で使える作成ルールとして開
発を行う非公式の共同グループである
GRADE Working Group によるGRADシ
ステムである。

一方で、その利欠点をみすえて、学会独
自の作成ルールを整備している団体もある
(利欠点を一覧表にしている)[AACE
2010][CMAJ 2007]。この中で、特にエビデ
ンスがない場合の取り扱いがGRADEシステ
ムと異なるのがthe U.S. Preventive Services
Task Force (以下USPSTF)らである(「After
deliberation, the USPSTF decided not to
adopt the GRADE approach.」と記載してい
るように独自の作成ルールである)[USPSTF
2009]。このように数ある作成ルールの中で
本稿は、各種ある作成ルールの中でも作成
ルールそのものの作成会議に参加が自由で
あるという、他の作成ルールに勝る利点があ
るGRADEシステムに従って、エビデンスがな
い場合の診療ガイドライン作成方法の、具体

的な手順を考察する。

さて、これまでにエビデンスがない場合、
「推奨なし」や「エビデンスが見つからなかつ
たので、委員会としては推奨文を作成しなかつ
た」、などの診療ガイドラインも多く存在す
る。もともとそれは、数あるクリニカルクエツ
ションの中には、このような場合もあるというス
タイルがほとんどである。しかし、本稿で扱いた
いのは、その診療ガイドラインで扱う疾患の
クリカルクエツションのほとんどが、このよう
な状況である場合である。もともと、そもそ
もエビデンスがない疾患に対して診療ガイド
ラインが必要かという問題もあるであろう。

また、本邦の診療ガイドラインでは、「エビ
デンスがないのでコンセンサスとして決定し
た」という場合も多い。しかし、そのほとんど
が手順が不明の状況で診療ガイドライン作
成委員会でいつの間にか決定されたもので
あることより、単なるコンセンサスレポートに
しか過ぎず、診療ガイドラインとしての役割が
満たされたとはいえない。本稿では、このよ
うなコンセンサスレポートにしかすぎないので
はなく、診療ガイドラインとして存在するた
めの具体的手順を示すのが目的である。たと
えコンセンサスでも利用者が誤解しないよ
うな配慮(資料作成)とコンセンサスにいた
った過程の公開が必須と考えている(だからこそ、
GRADEシステムのように、作成ルールの議

論に参加するのが自由であるシステムを推奨している)。なおPhillipsは、エビデンスが乏しい、またはない場合は、コンセンサスに基づく診療ガイドラインになるとしている[Phillips 2011]。

実際には、コンセンサスレポートに近くても診療ガイドラインと名称を付けている組織もあるので注意してほしい。たとえば、NCCE (National Comprehensive Cancer Network) では、表紙に、「Clinical Practice Guidelines」とあるにも関わらず、実際の内容で、エビデンスとコンセンサスによるカテゴリーとして、もっとも多い2Aなどが、「比較的低レベルのエビデンスに基づく推奨で、NCCNの統一したコンセンサスが存在する」とある[NCCE 2009]。

また、コンセンサスレポート以外にも、多くの資料の形態がある。たとえば、NICE (National Institute for Health and Clinical Excellence) のガイダンスには、診療ガイドラインを含めて以下の6タイプがある[NICE 2012](表3)。特に、5. Medical technologies guidanceと6. Diagnostics guidanceなどは、2010年以降と比較的最近に設定されたものである。よって、白岩らのNICEの紹介の記事には、記載されていない[白岩ら 2009]。これには、いずれも作成方法が決められており、それぞれ公開されている。しかし、いずれの資料も、すでにNICEの目的・組織や物事の決定の順序を理解している者のために書か

れたためであろうか、全体像を把握するのが困難な文章である[葛西 2011]。しかし、資料にはタイムラインが書かれており、たとえば、Guide to the single technology appraisal (STA) processの33ページなどをみると、作成までに39週(おおよそ1年)がかかることがわかる[STA 2006]。

そして、これらのガイダンスにおいて、医療資源の効率的利用を促進することを目的として設立されているNICE的なポイントとしては、安全性の検討は当然のことながら、費用効果の検討が必ず行われている点である。ここでNICEのガイダンスの種類を述べた理由は、これらの診療ガイドライン以外のガイダンスの定義をみることによって、診療ガイドラインがどのようなものであるべきか、そしてその作成の限界が明らかになると思われるからである。

新しい分野のために、いわゆるエビデンスが少ない場合、診療ガイドラインにこだわるのでなく、他のまとめも検討するべきであろう(もしかしたら、コンセンサスレポートになるかもしれない)。逆に成熟した分野であるにも関わらず、いわゆるエビデンスがない場合は、違うアプローチが必要になるかもしれないが、前述したように、早急に事例を集めて症例集積研究としてまとめ、エビデンスを作成する作業を、学会が怠ってはならないと考えている(早急に、ランダム比較試験を行うのは、現実的に不可能である)。

表3: NICEにおけるGuidance by type

-
1. Clinical guidelines: NICE clinical guidelines make recommendations to the NHS on treating and caring for people with specific diseases and conditions.
 2. Public health guidance: NICE public health guidance makes recommendations to the NHS, local authorities and other organisations in the public, private, voluntary and community sectors on how to improve people's health and prevent illness and disease.

3. Technology appraisal guidance: NICE technology appraisal guidance makes recommendations on when and how new and existing medicines and treatments should be used in the NHS.

4. Interventional procedures guidance: NICE interventional procedures guidance advises the NHS on whether new interventional procedures are safe and effective enough to be used routinely.

5. Medical technologies guidance: NICE medical technologies guidance helps the NHS to adopt medical technologies more rapidly and consistently by advising on efficacy and cost effectiveness.

6. Diagnostics guidance: NICE diagnostics guidance makes recommendations to the NHS on the efficacy and cost effectiveness of new diagnostic technologies.

いずれにしろ、ここでの議論は、「エビデンスがない場合」に推奨文を作るべきかであるが、実は、GRADEシステムは、基本的にエビデンスのない場合でも推奨文を作ることを推奨している(これに対しては、違う意見もあるので後述する)。その理由は、診療ガイドラインが患者や臨床医のためのものである限り、推奨文なしでは、利用者の混乱を招くだけであるので、推奨文の作成は必須であるとしている。

一方で、USPSTFは、検診などの健康な人への介入の場合、エビデンスが不十分の場合、推奨がない場合があるとしている。その場合は、recommendation推奨でなく、エビデンスが不十分であるという文章を勧告としている(The workshop resulted in a decision to transform what was formerly called an “I recommendation” into an “Is tatement,” as described elsewhere If a test or service existed but there was no evidence of net benefit, a decision not to offer the service is perfectly acceptable, because the patient is healthy and free of symptoms.)。そのためUSPSTFは、推奨がない場合を認めていないGRADEシステムを採用していない。しかし、Guyattは、GRADEシステムにおいても推奨がない場合があるため、そのような状況にも対応できるので、USPSTFの誤解だと記載している(「With GRADE ,guideline panels may choose to rate the quality of evidence as low or very low, and decide not to offer a

recommendation and to provide clarifying statements . Thus, GRADE accommodates the situation in which the USPSTF uses an I statement.)[Guyatt to the editor 2009]。たしかに、GRADEシステムを採用したSepsisの診療ガイドラインでは、no recommendationや、推奨文を作らないのと記載が存在する。そればかりか、この時に採用した投票であるGRADE gridでは、推奨なしの投票を認めている(この投票で一致しない時は、推奨文を作らないとしている)(Sepsisの診療ガイドラインには、Guyattも著者の一人に名前を連ねている:「The guidelines group was evenly split on the issue of SDD, with equal numbers weakly in favor and against recommending the use of SDD (Appendix H). The committee therefore chose not to make are commendation for the use of SDD specifically in severe sepsis at this time. The final consensus on use of SDD in severe sepsis was achieved at the last nominal committee meeting and subsequently approved by the entire committee (Appendix H provides the committee vote).」)[Sepsis 2008]。

しかし著者は、このGuyattの文章は、これまでのGRADEシステムの理解と違うと考えている。そこで、GRADEシステムの広報を担当しているYngveFalck-Ytterに私信で質問した内容をYngveFalck-Ytterの許可を得てBoxに記載したので参考にして欲しい。著者は、

GRADEシステムは、YngveFalck-Ytterと同様に基本的に推奨なしを認めていないと理解している。

それ以外にも、たとえばAACE (American Association of Clinical Endocrinologists) は、エビデンスがない場合 (No evidence) を、「theory, opinion, consensus, or review」と定義し、single case report すらない場合のみを No evidence としており、本稿の定義と異なる。エビデンスがなく、推奨が4段階の4番目と考えられる場合でも他の要因によっては推奨が4段階の3番目に上昇する場合もあるとす

る作成ルールである。すなわち、具体的な手順が明らかになっていないことが問題であるものの、エビデンスがない場合で、GRADEシステムと同様に推奨度を決定するシステムがあるので診療ガイドライン作成委員会ならば、これらの作成ルールなども勉強して各種作成ルールに精通する必要がある。

このように、エビデンスがない場合に推奨文を作るか否かに関しては、いろいろな議論があるが、本稿では推奨文を作ることを前提として議論を進める。

Box: YngveFalck-Ytterのレター

著者よりYngveFalck-Ytterへの質問

Dear Yngve, I want you to teach me. In dentistry, there is not often evidence. In Cochrane Oral Health Group, there are a lot of results "no evidence".
<http://onlinelibrary.wiley.com/o/cochrane/clsystrev/articles/CD006541/frame.html>
There is not the cohort study in observational study, too. Evidence is only case series study with few case or case reports.
How do we think about a recommendation for CQ of "no evidence"? Or we make recommendation of "no recommendation" or "no evidence"?
I read the following articles. The American Society of Hematology 2011 evidence-based practice guideline for immune thrombocytopenia
<http://bloodjournal.hematologylibrary.org/content/117/16/4190.full.html>
However, my question did not solve. I want your advice.
Thank you. Hidemichi Yuasa

YngveFalck-Ytterより著者への回答

Dear Hidemichi. Good to hear from you again. I hear this very often - there is "no evidence". Well, most of the time I hear this what was actually meant was: "There is no good (or high quality) evidence". Complete absence of evidence is actually quite rare. I would even argue:
If there is a question which intervention to recommend (out of at least 2 competing management strategies), then there ought to be evidence, because otherwise we would not have the question!
The nice thing about GRADE is, that you are not limited to the source of the evidence. Everything counts: from RCTs to a range of observational studies (e.g., cohort,

case-control, case series etc.) to in vitro studies to lab studies etc. Obviously, our confidence in the estimate of effect will be lower (and evidence from case reports or case series likely going to be very low quality).

However, I don't think it is in the best interest of clinicians and patients if guideline developers state:

"There is insufficient evidence and therefore we cannot make any recommendation". The USPSTF does that quite a bit. And again, I don't think this is helpful because clinicians (as do patients) have to make a decision what to do. They don't have the luxury to say "OK, the evidence is insufficient, then I don't do anything..."

4. 一般的な推奨文を作るための具体的な手順

さて、エビデンスがない場合に推奨文を作るための具体的な手順を考える前に、一般的な推奨文を作るための具体的な手順を確認しなければならない。本稿では、GRADEシステムが基本であるが、一般的な診療ガイドライン作成の手順(ルール)を表4にまとめた。これをさらに要約すると、(1) エビデンスの質・効果の大きさ・不利益に関する系統的な資料を作る、(2) その資料を使って、診療ガイドラインパネル会議で推奨文を作成する、の2点となる。この2点を明確に区別することが、エビデンスがないと言って作業が中断しないための、最も重要なポイントだと考えている(図1)。

この重要なポイントは、多くの作成ルールで明確に区別されていない。GRADEシステムでも、多くの解説論文を読んでもすぐに理解できない点であろう。その理由は、海外では、もともと系統的レビュー担当者や図書館司書などが整備されている環境で考えると、当然のこととして受け止められているからである(NICEの診療ガイドラインのページを参

照しても、これらの順番や組織編制を、まず理解しないと、各論を学んでも意味がわからないままであると言える)[NICE手順 2012]。たとえば、相原のWebサイトなどにあるスライドや、Institute of medicineの資料などを見れば当然のこととして区別されているのが理解できると考えられる[相原 2011][Institute of medicine 2011]。しかし本邦では、この区別を理解している者が少ないのが現状である。

この診療ガイドラインパネル会議は、パネリストの選択が、選択に対する客観性が確保されてないという問題や、科学的・系統的に作成した資料をもとにするが最終的に個人の判断によって決定されることが科学的でないとの批判もある。しかし、これらの資料をもとに議論し、これらの資料に立ち返る訓練が、パネリストに行われていれば問題ないであろう。また、このような診療ガイドラインパネル会議が行われるようになったのは、患者の不利益とのバランスを考慮することなくエビデンスの質のみで自動的に推奨の強さが決定するシステムの反省である(この、権威者の思惑が入らない自動的に決定するシステムには、当時著者も感銘を受けた一人である)(表5)。

表4: EBMに最適化した推奨を作るための具体的な手順

I. 診療ガイドライン作成委員会

1. クリニカルクエッチョンを収集し、診療ガイドラインの目標をたて、計画書を作る
EBMステップ1

II. 系統的レビュー担当(専門スタッフ)

2. エビデンスの検索・収集・分別
EBMステップ2 (EBMの検索と系統的レビューの検索では戦略が異なることに注意)
3. アウトカムごとのエビデンスの質を評価し、アウトカムごとのエビデンスの効果の大きさをまとめ資料とする(エビデンスの質と効果の大きさ)

EBMステップ3

4. 害や負担やコストなどに関する資料を作成する (EBMステップ3)

III. 診療ガイドラインパネル会議担当者(パネリスト)

5. その資料をもとに診療ガイドラインパネル会議で検討し、そのクリニカクエッチョン全体のエビデンスの質と推奨度(確信の強さと方向)を決定して推奨文を作成する

EBMステップ4

IV. 診療ガイドライン作成委員会

6. 診療ガイドラインの本文やパンフレットなどを作成し配布する

EBMステップ5

図1: 診療ガイドラインパネル会議とそのために作る資料の関係について

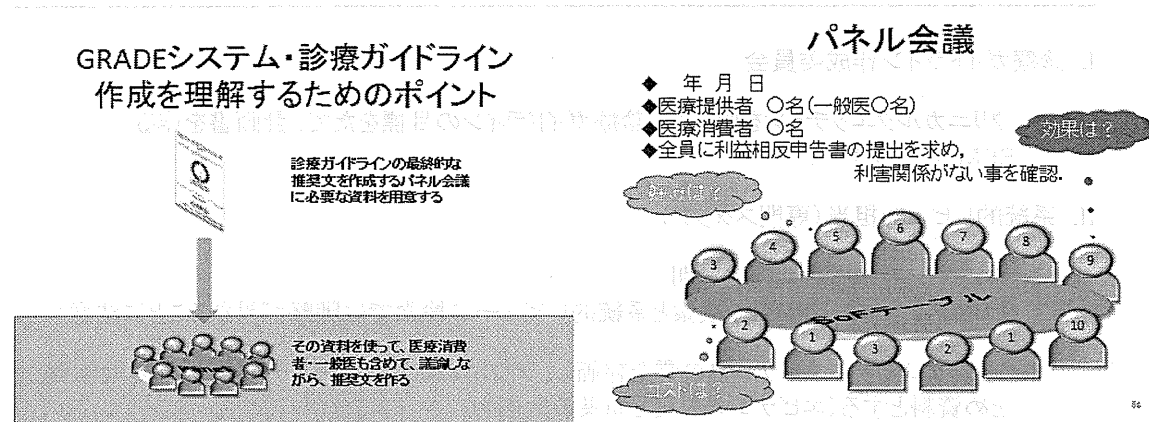


表5: 診療ガイドライン推奨の決定方法の変遷

権威者による密室内での推奨決定



論文の研究デザイン(アウトカムは何でも良い)などで、権威者や利益享受者の思惑がないらないように自動的に推奨の強さが決定



医療消費者も含めたオープンな診療ガイドラインパネル会議(経済・法律にも詳しいものの参加があればさらに良い)で、あくまでもエビデンスに基づきながら利益と不利益のバランスを考慮して推奨を決定

すなわち、まず行わなくてはならないことは、各アウトカムに対するエビデンスの質の評価とエビデンスの効果の大きさの資料を、どのように作成するかを考えることであり、その具体的な手順が理解できないといけないということである(表6の(B)の資料作り)。この作業は、GRADEシステムというより、系統的レビュー担当の者が行うのでコクランレビューで使われるコクランハンドブックV5に従うことになる[コクランハンドブックV5]。ここで注

意が必要なことは、前述したようにGRADEシステム・コクランレビューでは、エビデンスの質と各論文にバイアスのリスクが多いか少ないかを厳密に区別しているが、他の作成ルールでは、区別がなかったり、各論文のバイアスのリスクをエビデンスの質と表現したり、各論文の研究デザインの種類をエビデンスの質と表現しているので注意がいる。

さてここまで理解すると、今度は、(1) 各ア

アウトカムに対してのエビデンスの質の評価(この中に、表6の(A)各論文の限界・バイアスのリスクの評価と、表6の(B)その評価を含めて論文間の一致性などを評価した、エビデンスの質の評価がある)(また、表6の(C)その後の各アウトカムを総合的に評価して、そのクリ

ニカルクエッションのエビデンスとしての質の評価は、診療ガイドラインパネル会議で行うので注意すること)と、(2) 各アウトカムに対してのエビデンスの効果の大きさの評価について、より詳細な手順を知る必要がある(これ以降は、この2つを厳密に区別する)。

表 6: コクランハンドブック V5 によるエビデンスの質と結果の大きさのまとめについて[湯浅 2011]

-
- (A) その疾患の治療に対するアウトカムごとに作業を行う
一つ一つの研究結果(個々の論文)ごとに作業を行う(アウトカムごとに一つの論文を何度も検討)
→バイアスのリスクを検討
→結果の大きさを検討
- (B) その疾患の治療に対するアウトカムごとに作業を行い、資料を作成する
(A)で評価したすべての研究結果(複数の論文)をまとめて作業を行う
→研究間のばらつきなどを検討
→エビデンスの質を評価
→結果の大きさを評価
- (C) その疾患の治療ごとに(B)で作られた資料を診療ガイドラインパネル会議で検討
(B)で評価したアウトカムごとの研究結果のまとめを、さらにまとめて作業を行う
→アウトカムの相対的なバランスを評価
→全体的なエビデンスの質を評価
→全体的なエビデンスの結果の大きさを評価
- (D) 害、患者の価値観や好み、コストや資源についての資料を作成する(他の資料)
-

・GRADEシステムでは、各論文を一つの研究単位(論文単位)とするのではなく、各論文の中のアウトカム一つ一つを、一つの研究単位(アウトカム単位)として考えていることに注意すること。

・エビデンスの質の評価と効果の大きさの評価と他の資料作りの3つの作業がある。

・(C)は、系統的レビュー作成グループが行うのではなく、診療ガイドラインパネル会議で決定される

4-1. 各アウトカムに対してのエビデンスの質の評価について

多くの作成ルールで、ランダム比較試験の研究デザインを持つ論文一つの研究の限界(表6の(A))や、複数の論文によるエビデンスの質を決定するための具体的手順は詳細に述べられている(表6の(B)または(C))。しかし、ランダム比較試験以外の論文しかない場合の具体的手順が述べられていることは少ない。しかも、あったとしても、その多くが、ある程度のバイアスが少なく、計画された観察研究がある場合であり、いわゆる症例集積研究などの場合はほとんどない。そこで、まずランダム比較試験と観察研究のどちらが重要かという問題(現在の一般的な認識ではランダム比較試験のが重要と考える者が多いと思うが歴史的には多くの議論がなされている)と観察研究に対するエビデンスの質の評価を確認する必要がある。

4-1-1. ランダム比較試験と観察研究のどちらが重要かという問題について

質の高いエビデンスが少ない場合を考えると、これまでランダム比較試験と観察研究とを比較した多くの議論を理解することが必要なので、簡単に流れを示す(実際には、議論はさらに複雑で、時系列でないが、全体像を掴むため簡単にまとめた)。

まず、(1) ランダム比較試験の整備が進められた、(2) 観察研究よりランダム比較試験が、効果の差の検出が正しいとされる論文が増えた、(3) メタ分析・系統的レビューには、質の低い論文は除外する、(4) ランダム比較試験でも、大規模臨床試験が、効果の差が正しとされる、(5) メタ分析・系統的レビューには、ランダム比較試験だけでなく、観察研究も含めるべき、(6) ランダム比較試験と観察研究の効果の差の推定が異なり、どちらかという観察研究のが正しいのではないか、(7) ランダム比較試験後に観察研究を行っ

て確認すべきであるという流れで議論されてきた。

最後の(7)は、あたりまえのことであり、以前から言われていたにもかかわらず、観察研究がランダム比較試験よりもレベルが低いとされていたため、行われることが少なかった。またスポンサーも、せっかくランダム比較試験で得られた結果に対して、わざわざ確認するよりも、ランダム比較試験で得られた結果を宣伝することに、資金を投入するためであった(これは、製薬会社のみならず公的資金を獲得した研究者も次の資金獲得のための努力を優先する)。

すなわち注意が必要なことは、以上の議論は、製薬会社などのスポンサーが多く、ランダム比較試験を行いやすい内科系の場合であるということである。ランダム比較試験そのものが困難な外科系や、資金が乏しい歯科分野では、このような研究デザインの議論すら少ないのが現状である(数百万円の資金も獲得できない場合の研究デザインの整備も現実的に必要にもかかわらず行われておらず、数千万円と数億円の、ランダム比較試験の小規模と大規模の違いのみの議論がほとんどであった)。このように記載すると、「いや、私の医局では、低予算で質の高いランダム比較試験を実施している」との批判があるかも知れないが、一般的には不可能である。

以上の流れを理解したうえで、まず観察研究のバイアスのリスクの評価について説明する。このことが、「エビデンスがない場合」の質の評価につながると考えられる。

4-1-2. 観察研究のバイアスのリスク(論文の質)の評価について

これまで多くの評価がなされていたが、もつとも有名な観察研究の質を評価するのがNewcastle-Ottawa Scaleなどである(本稿は、観察研究を主とするのではないので、具体

的なスコアの記載は行わないが、いろいろなチェックリストをまとめたものが、GRADEシステムの解説論文に記載されているので参考にして欲しい)[GRADE limitations 2011]。注意が必要なことは、この評価は、あくまでもランダム比較試験におけるJadadスコアや、コクランハンドブックにおけるバイアスのリスクと同じように個々の論文のデザイン上の質を評価しているにすぎない(表6の(A)でもない)[Jadad 1996]。すなわち、その知りたいアウトカムに対するエビデンスの質(表6の(B))や、クリニカルクエッションに対する回答となるエビデンス全体として考えた場合の質(表6の(C))ではない。よって、GRADEシステムのように、質の高い観察研究が複数あっても、その結果が不一致の場合、全体のエビデンスの質は低くなるべきであるが、それらの方法論は確立されていない。

これらの報告されている観察研究の質を評価する方法を学習することは重要だが、一方で、これら観察研究の質の向上のために作られた観察研究を行う場合のチェックリストを理解することも、観察研究によるエビデンスの質を評価するのに重要であろう(STROBE 声明・MOOSE チェックリスト)[STROBE 2007]。

4-2. 各アウトカムに対してのエビデンスの効果の大きさの評価について

次に、各アウトカムに対してのエビデンスの効果の大きさの評価(表6の(B)であり、(C)は診療ガイドラインパネル会議で決定するもので別)についてだが、これも、ランダム比較試験と観察研究での整備は進んでいるものの(観察研究では、まだ整備されているとは言えないが、少なくともその方向に進んでいる)、いわゆる症例集積研究・症例報告に対しての整備は行われていない。

ランダム比較試験と観察研究においては、いわゆるメタ分析という統計学的手法が各種

開発されている。これらの統計学的手法は、単に効果の大きさをまとめるだけでなく、結果の一致性などを測る方法も開発されている。詳細は、コクランハンドブックV5などを参考にして欲しい。しかし、観察研究に対しては、まだ開発途中でもあり、たとえば2011年のコクラン共同計画の集会(コクランコロキアム)でも、複数の方法論の報告があった[Cochrane Collaboration 2011]。

5. エビデンスがない場合の推奨文を作るための具体的な手順

さて、以上で一般的な推奨文を作るために必要な(1) 各アウトカムに対するエビデンスの質の評価(表6の(B))と(2) エビデンスの効果の大きさの資料の作成ルールを確認した。次に、「エビデンスがない場合」、すなわち「エビデンスが症例数の少ない症例集積研究か症例報告しかない場合」に、それから評価したエビデンスのみで診療ガイドラインパネル会議で使用する、各アウトカムに対するエビデンスの質の評価とエビデンスの効果の大きさの資料を、具体的にどのように作成するのかを考えることとする。この資料は、系統的レビューを、症例数の少ない症例集積研究か症例報告で作成することとほぼ同じである(表6の(B)の資料)。これを数量的なメタ分析まで可能かの議論も含めて考える必要があるが、ここでは、Jenicekの「症例の質的メタ分析を探求し発展させなければならない」という言葉を紹介することにとどめる[Jenicek 2002]。

その理由は、ここでいくら議論を展開しても、それは私見にかすぎず、科学的に妥当性の検証などを行ってないからである。そもそも洗練されたコクランレビューのためのコクランハンドブックに記載されている各論文のバイアスのリスクを検討するチェックリストですら妥当性の検討がなされていない。若干古い情報だがJadadによると、数多い、エビデンス・論文の質を評価するスコアなどで、唯一

妥当性が評価されているのが、一つの論文の質を評価するJadadスコアであるとしている[Jadad 2004]。よって、本稿で提示する具体的な手順(作成ルール)の案に対しても、今後の議論のためのきっかけにしかすぎず、より多数で公開された議論が必要と考えている。著者は、各学会が独自で作成ルールを作り乱立することは、利用者の混乱をまねくだけと考えており、このような案を紹介することに抵抗がある。しかし、現時点で困っている診療ガイドライン委員会も多いと思っている。今後これらの場面にも対応が可能なようにGRADEシステムが整備されれば、本稿の案はいらない(現時点でGRADEシステムは、これらの場面にも対応されているとの意見もあるが、実際は、具体的な手順が示されておらず、どのように作成していけばよいか不明である)。

実際の具体的な手順は、一般的な場合と同様に、(1) 各アウトカムに対してのエビデンスの質の評価と、(2) 各アウトカムに対してのエビデンスの効果の大きさの評価について、わけて考える必要がある。

ここで(1)の各アウトカムに対してのエビデンスの質の評価の中のいわゆる論文の限界・バイアスのリスク(表6の(A))に参考になるのは、症例報告を読むためのチェックリスト[Atkins 2002]や症例報告を書くためのチェックリストである[Huston 1996]。しかし違いもあり、これらの症例報告は、新しいか稀少な症例が良いとされていることである。診療ガイド

ラインでは、稀少な症例でなく、より一般的な状況を前提としている。よって、多くの専門医にコンセンサスは得られており、すでにルーチン化されている場合などの場合、コンセンサスレポートは存在し、確かに症例があるものの、学術的な症例報告になってない場合もある。この場合は、診療ガイドライン作成委員会のメンバーが、上記のチェックリストを基に、科学的な症例報告を記載する必要がある。この場合、もっとも参考になるのが、Jenickの教科書である[Jenicek 2002]。

本稿では、以下の「案」を提示したが(表8)、その本質は、EBMの理解の基本的な概念である外部妥当性(一般化可能性)の考え方(内部妥当性は、いわゆる論文の限界・バイアスのリスク(表6の(A))に近いと考えられる)や因果関係の妥当性(criteria of causal association)を理解することが必要と考えている[湯浅 1998][Yuasa 2002][津田 2001][津田 2009](表6の(B)と(C)ならびに推奨文そのもの)にも相当するので、区別が困難である)。これらの考えを十分に理解しない限り、表面上のチェックリストのみ作成しても意味がない。

さらに誤解があるのが、EBMの批判的吟味という用語であり、これを「論文にケチをつけること」と考えている者が多数いることである。EBMの批判的吟味は2つの目的があり、その中の一つを系統的レビューでも目的としている(表7)。

表7: EBMと系統的レビューの批判的吟味・バイアスのリスクの検討の意味

EBMにおける批判的吟味:

- (1) 忙しい臨床の現場で、読まなくて良い論文をいち早くみつけるため
- (2) バイアスを乗り越えて使えるデータ(宝物)を探すため

系統的レビューのバイアスのリスクの検討:

(1) バイアスを乗り越えて使えるデータ(宝物)を探すため

次に、外部妥当性について少し記載すると、Campbellらは、1963年では、「External validity, of course, deals with the issue of generalizability of the results found to other populations, settings, and so forth」としているが)、1979年では、「external validity: They have to do with (1) generalizing to particular target persons, settings, and times, and (2) generalizing across types of persons, settings, and times.」と、「to」から「across」への変化させているように、医療消費者も含めた議論により、より幅広い場面で使える診療ガイドラインを目指すべきであろう(海外では、これらの外部妥当性や因果関係に対する議論を学生の間より十分に教育されているので、そもそも本邦とは土俵が異なる)[Campbell 2000]。

さらに、因果関係についても記載する。妥当性を検証する基準は多くあるが、有名なものはKochの3原則、Evansの8条件、Hillの9基準(視点)等であるが、Wardの論文の参考文献にリンク先が整理されているので利用できる(疫学での因果関係と、基礎実験での因果関係と、一般の社会での因果関係は、概念が違うことに注意されたい(最近では、臨床試験や医学統計などで因果推論という用語も使われるが、これも含めて立場が異なる)) (著者は、ある統計学的な研究デザインのみで因果関係が推論できるという文章は、誤りであるという立場である)[Ward 2009]。この因果関係を各研究から考えてみるのが、エビデンスの質の評価に役立つと考えてい

る。すなわち、もともとエビデンスの質そのものが、因果関係が強いかわかりの評価と類似点が多いからである。しかし、これらの症例数の少ない症例集積研究か症例報告だけで因果関係が言えるかどうかということは、本稿の主旨ではない。しかし著者は、因果関係を考えるのは重要だが、必ずしも因果関係がなければ治療を行ってはいならないと言うことではないと考えている(このような因果関係についても診療ガイドラインのパネリストは学ぶ必要がある)。

5-1. 各アウトカムに対してのエビデンスの質の評価について

Jenicekの教科書などを参考に、各アウトカムにおける個々の論文の評価(限界・いわゆる論文の限界・コクランハンドブックでいうところのバイアスのリスクの評価に相当)するチェックリストを作成した(表6の(A)) [Jenicek 2002]。もちろんチェックリストの妥当性の検討は行っておらず、チェックで何個が「はい」の場合に論文が良いと評価するかのカットオフポイントも示していない。しかし、これらの多くが「はい」となれば(「いいえ」と「不明」が少ない)、その論文の評価は高いと考えることに、多くの方が同意されると思う。

特に、「症例選択基準は明確か:都合の悪い症例を隠している可能性はないか」が重要であり、このことは、たった一つの症例報告からでは不明かもしれないが、症例集積研究だと評価可能であろう[湯浅 2004]。

表8: 症例集積研究・症例研究から得られた一つのアウトカムに対する論文・研究の評価のチェックリスト(案)

症例選択基準は明確か: 都合の悪い症例を隠している可能性はないか

診断基準は明確か: 本当に治療が必要な症例を選んでいるか

評価基準は明確か: 評価の基準が都合のいいように変わっていないか

症例のリスクは示されているか: 診断・予後に関わる他のリスクが評価されているか

予後は長期か: 短期予後がよくても長期予後がかえって悪くないか

症例・治療は新しいか稀なものか: 報告に臨床的意義があるか

有害事象が記載されているか: いわゆる副作用や合併症がないか

・「はい」「いいえ」「不明」で検討

・なお、「症例選択基準は明確か: 都合の悪い症例を隠している可能性はないか」が「いいえ」の場合は、それだけで評価は低いとする。

次に、個々の症例数の少ない症例集積研究か症例報告から得られた各アウトカムのエビデンスの質の評価(表6の(B))だが、基本的な考えは、GRADEシステムの観察研究の場合と同じである(エビデンスの質がVery Low(非常に低)から各論文間の関連を考えてレート・アップ(→低→中→高)していく)。しかし、GRADEシステムでは、このレート・アップの場合に、各論文の評価(いわゆるバイア

スのリスク)が考慮されていない。レート・アップしてから、研究の限界(いわゆるバイアスのリスク・表6の(A))でレート・ダウンするのかなど明らかとなっていないのが現状である(この議論は、GRADEシステムの手順をしっかりと理解しないと意味不明と考える)。よって、本稿では、レート・アップの前提条件として、各論文の評価(研究の限界、いわゆるバイアスのリスク)を組み込むこととした。

表9: 症例数の少ない症例集積研究か症例報告から得られた一つのアウトカムに対する複数の論文・研究からのエビデンスの質の評価におけるレート・アップ(案)

(1) 症例数の少ない症例集積研究か症例報告から得られた各アウトカムのエビデンスの質は、Very Low(非常に低)から開始する。

(2) 「症例集積研究・症例研究における各論文の評価のチェックリスト(案)」で「はい」が多くて良いと評価できる論文がほとんどであることが必要。もし少なければ、以下の(3)(4)がプラス(レート・アップ)

であってもエビデンスの質はVery Low(非常に低)とし、その問題点を列挙して診療ガイドラインパネル会議の資料とすること。

(3) 関連性 (association) の強さ・関連性を示す非常に有力なエビデンス: 治療などの介入の推定効果が非常に大きく、妥当性を脅かすものがない直接的エビデンスに基づいており、2つあるいはそれ以上の報告から得られた一貫性のあるもので、ありそうな交絡因子がない場合。(+1)

(4) 用量反応勾配のエビデンス: たとえば、ワーファリン治療例でINR値が高くなるほど、出血が多くなることが各症例報告により、高い確信となっている。(+1)

・評価は、4段階で「very low 非常に低」・「low 低」・「moderate 中等度」・「high 高」

・「+1」は、2つのため、エビデンスの質がHigh (高)にはならない。

5-2. 各アウトカムに対してのエビデンスの効果の大きさの評価について

多数の症例報告の集まりに基づく「メタ分析」が可能ならば良いが[Cook 1996]、現時点で、その方法論は確立されておらず、Jenicekが希望しているような議論もほとんど進んでいない[Jenicek 2002]。数少ない報告として、Robinsonらは、症例集積研究からロジスティック回帰分析を利用したメタ分析を行っている(症例集積研究の症例数は、最低が6症例の研究で、最高が2122症例の研究であった)[Robinson 1997]。他には、Prostate Cancerの診療ガイドラインでは、選択した436論文のなかで、症例集積と症例報告が353論文で、比較試験が28論文であり、グラフ化することによって、資料を作成しているが、完全なものでないとの注意書きがある

(These graphs are an intermediate work product. As such, they are subject to a number of problems such as possible mistaken data, redundant data groups separated by factors irrelevant to the graph. Thus, the Panel considered these graphs to be sufficiently heterogeneous so that conclusions could not be drawn based on the data and that further refinement would not be helpful.)[AUAER 2007]。また、Sheaらは、症例報告を集めて死亡率や合併症の発生率などを推定しようとしている[Shea 1996]。

以上より、現時点で推奨できる方法はないが、その疾患分野で工夫することが大切と考えている。本稿では、特に重要と考えられるポイントを表としたので参考にして、診療ガイドラインパネリストが見やすく利用しやすい資料を作成して欲しい。

表10: エビデンスの効果の大きさの評価をまとめる場合のポイント

- ・その論文の当該施設での、その症例の受診率がどのくらいで、集積期間を記載する
- ・年齢・性別・人種・重症度・治療期間などで分類する(細分類しすぎないこと・分類が交絡因子となる場合は、平均値などの代表値を算出するに留める)
- ・研究ごとに本当に治療法の違いがないかの確認する、特に治療に用量があるならば記載すること
- ・奏功までの期間と奏功の有無をカウントする(生存期間なら、その期間を一覧として記載)(アウトカムが脳梗塞の発生などのイベントの発生ならば、その有無とイベント発生までの期間となる)
- ・合併症・害について記載があれば、一覧とする

- ・症例報告以外にも、コンセンサスレポートなどがあれば、製作者や団体のスポンサーを明示した上で添付資料とする。
- ・上記項目を一覧表にするのか、研究ごとに一枚のカードにまとめるかなど、診療ガイドラインパネルリストが見やすく利用しやすい資料を作成すること。

6. 資料を診療ガイドラインパネル会議に提出する

資料がそろった後の診療ガイドラインパネル会議での議論は、基本的にGRADEシステムに従って良いと思われる。資料さえあれば、利益と不利益のバランスを議論しまとめることになるため、経験のある診療ガイドラインパネル会議のメンバーなら作業は進行するはずである。経験が少ないならば、臨床決断に関する文献を読んで訓練する必要があるだろう[JAMA 2010][Gross 2002]。ここでパネルリストが十分な訓練を受けてないと、せっかく系統的レビューなどによって作成したエビデンスから離れて、個人の経験や思い(大切な要因ではあるが)のみで議論してしまうことになる。

しかし、若干の注意も必要と考えられる。

たとえば、診療ガイドラインパネル会議の構成人数としては、エビデンスの質が、低(low)なので、その介入に対して客観的に判断できる者を多くし、その会議参加者の背景を明確にすることが重要である。客観的に判断できる者の人数配分などを具体的に提示することはできないが、少なくとも、その治療の推進者以外の専門家が複数いることが必要だろう(決して、推進者と反対者が同数になる必要があるということでないし、反対者が多くなければならないということでもないし、少数派の意見を聞くことが露骨な誘導でもない)。そして、意見がある程度収束した場合でも全員一致とならない限り投票結果を公表すべきと考えている。また、本邦では費用対効果分析が乏しいため、これらの分野に精通した者の参加も必要である。

そして、たとえ推奨が条件付きとなっても

問題はないが、少なくとも、安易に「推奨文を作成しない」に逃げずに、責任を持つことが社会にとって必要であることを診療ガイドラインパネル会議の冒頭で十分に議論し意見を統一することが必要である。このようなプロセスを踏んでも、なお推奨文を作成しないとの意見となった場合、そのことを詳細に公開する必要がある(少なくとも臨床医に、その判断をゆだねる資料がなければ、資料の価値がない)。そして、それに対して学会での責務が何であるか(たとえば、臨床試験を実施するなど)を具体的に示す必要がある。

7. 最後に

エビデンスがない場合であっても、社会がどうしてもその分野の診療ガイドラインを必要としているのならば、診療ガイドラインを作成するという困難な問題に立ち向かうのは大切なことである。そこで、その様な困難に立ち向かう、診療ガイドライン作成委員ならびに、系統的レビュー担当者や診療ガイドラインパネリストに、Jenicekの名文を送りたい。それは、「EBMを実践する者は、しばしばエビデンスが見あたらなかつたり、あってもその質や完全性がたびたび欠落していたりしていることに欲求不満を感じてはいけな。ここに医学の難しい問題の一つ、つまり不確実性の中で、そして情報やエビデンスが不完全な状況で正しい判断を下すこと、が横たわっているのである。それはまた医学の美であり、それを正しく実践する者に対するとてつもない報酬なのである。」という、臨床医として大切な心構えである[Jenicek 2002]。

参考文献

[AACE 2010] Mechanick JI, Camacho PM, Cobin RH, Garber AJ, Garber JR, Gharib H, Petak SM, Rodbard HW, Trencle DL; American Association of Clinical Endocrinologists. American Association of Clinical Endocrinologists Protocol for Standardized Production of Clinical Practice

Guidelines--2010 update. *EndocrPract.* 2010 Mar-Apr; 16(2):270-83.

[AHA 2010] American Heart Association. *Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines.* 2010.

[相原 2010] 相原守夫、他: 診療ガイドラインのための GRADE システム(初版). 凸版メディア. 2010.

[Atkins 2002] Atkins, Clare and Sampson, Jennifer, "Critical Appraisal Guidelines for Single Case Study Research" (2002). ECIS 2002 Proceedings. (<http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1011&context=ecis2002> 2011/8/19アクセス)

[相原 2011] 相原内科. GRADEシステム. GRADE flow chart. (<http://www.grade-jpn.com/> 2011/8/23アクセス)

[AUAER 2007] American Urological Association Education and Research. *Guideline for the Management of Clinically Localized Prostate. Cancer: 2007 Update. Appendixes* (<http://www.auanet.org/content/guidelines-and-quality-care/clinical-guidelines/main-reports/proscan07/appendixes.pdf> 2011/8/19アクセス)

[Campbell2000] *Validity and Social Experimentation Donald Campbell's Legacy Volume: 1* Editor: Leonard Bickman : Vanderbilt University, Nashville Published By: SAGE Inc.2000.(原著より入手しやすいので参考文献としたがtoとacrossは、原著でも確認)

[CMAJ 2007] Palda VA, Davis D, Goldman J. A guide to the Canadian Medical Association handbook on clinical practice guidelines. *CMAJ.* 2007 Nov 6; 177(10):1221-6.

- [Cochrane Collaboration 2011] the 19th Cochrane Colloquium. P2B245: Does stratification analysis indicate the confounding factors in meta-analysis of observational studies? Naing C, Mak JW. P3B372: The hazards of extracting hazard ratios from survival curves for meta-analyses in observational studies. Imamura M, MacLennan S, Dahm P, Neuberger MM, Omar MI, MacLennan G. (<http://2011.colloquium.cochrane.org/> 2012年2月13日アクセス)
- [コクランハンドブック V5] Julian P. T. Higgins, Sally Green. Cochrane Handbook for Systematic Reviews of Interventions (1st). Wiley. 2008.
- [Cook 1996] Cook MW, Levin LA, Joseph MP, Pinczower EF. Traumatic optic neuropathy. A meta-analysis. Arch Otolaryngol Head Neck Surg. 1996 Apr; 122(4): 389-92.
- [GRADE limitations 2011] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol. 2011 Apr; 64(4):407-15. Epub 2011 Jan 19.
- [Gross 2002] Richard Gross. 臨床決断のエッセンス 不確実な臨床現場で最善の選択をするために 今井裕一訳 医学書院. 2002.
- [Guyatt to the editor 2009][Petitti in response 2009] Guyatt GH, Helfand M, Kunz R. Comparing the USPSTF and GRADE approaches to recommendations. Ann Intern Med. 2009 Sep 1; 151(5):363; author reply 363-4.
- [Hahn 2006] Hahn T, Jones R, Wall D. ASBMT Evidence-Based Reviews for Pediatric and Adult Acute Lymphoblastic Leukemia. Hematology, Oncology, and Immunology. 16(1). 4-9. 2006. (http://static.cjp.com/gems/mtr/BMTR_16.1.pdf 2011/8/17 アクセス: The complete papers can be found in Volume 11, Issue 5 (pages 823-861), and Volume 12, Issue 1 (pages 1-30) in Biology of Blood and Marrow Transplantation.)
- [Huston 1996] Patricia Huston,; Bruce P. Squires. Case reports: information for authors and peer reviewers. Canadian Medical Association Journal 1996; 154: 43-45. (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1488107/> 2011/8/19アクセス)
- [Institute of medicine 2011] 米国医学研究所. Clinical Practice Guidelines We Can Trust. (<http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx> 2012年2月13日アクセス)
- [JAMA 2010] 相原 守夫、他 監訳: 医学文献ユーザーズガイド 根拠に基づく診療のマニュアル(第2版). 凸版メディア. 2010.
- [Jenicek 2002] Milos Jenicek. EBM時代の症例報告. 医学書院. 2002. (新刊入手困難のため、カーリル (<http://calil.jp/book/4260138855>) で検索が良い).
- [Jadad 1996] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996; 17: 1-12
- [Jadad 2004] Alejandro R. Jadad (アレハンドロハダット (著)). ランダム化比較試験—100のquestionsに学ぶ. じほう. 2004.
- [葛西 2011] 葛西美恵. 英国における【経済評価の政策的利用】. medicina. 48巻(10号). 2011. (<http://www.igaku-shoin.co.jp/misc/medici>

na/shohou4810/ 2012年1月22日アクセス)

[NCCE 2009]NCCE. NCCNガイドライン.
(<http://www.tri-kobe.org/nccn/index.html> ・
http://www.nccn.org/professionals/physician_gls/categories_of_consensus.asp 2012年1月22日アクセス)

[NICE 2012] NICEにおけるGuidance by type. 2012.
(<http://www.nice.org.uk/guidance/index.jsp?action=byType> 2012年1月22日アクセス)

[NICE手順 2012] Developing NICE clinical guidelines. 2012.
(http://www.nice.org.uk/aboutnice/howwe-work/developingniceclinicalguidelines/developing_nice_clinical_guidelines.jsp 2012年1月22日アクセス)

[Phillips 2011] Sue Phillips. Formulating recommendations in the absence of evidence: time for more rigorous methods. 8th G-I-N conference (Guidelines International Network Conference 2011).

(http://www.g-i-n.net/events/document-store/g-i-n-conferences/seoul-2011/presentations-seoul-2011/short-oral-seoul-2011/o38_phillips_formulating-recommendations-in-absence-of-evidence_290811.pdf 2012年2月13日アクセス)

[Robinson 1997] Robinson JW, Dufour MS, Fung TS. Erectile functioning of men treated for prostate carcinoma. *Cancer*. 1997 Feb 1;79(3):538-44.

[Sepsis 2008] Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, Reinhart K, Angus DC, Brun-Buisson C, Beale R, Calandra T, Dhainaut JF, Gerlach H, Harvey M, Marini JJ, Marshall J, Ranieri M, Ramsay G, Sevransky J, Thompson BT, Townsend S, Vender JS, Zimmerman JL, Vincent JL; International Surviving Sepsis Campaign Guidelines Committee; American

Association of Critical-Care Nurses; American College of Chest Physicians; American College of Emergency Physicians; Canadian Critical Care Society; European Society of Clinical Microbiology and Infectious Diseases; European Society of Intensive Care Medicine; European Respiratory Society; International Sepsis Forum; Japanese Association for Acute Medicine; Japanese Society of Intensive Care Medicine; Society of Critical Care Medicine; Society of Hospital Medicine; Surgical Infection Society; World Federation of Societies of Intensive and Critical Care Medicine. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med*. 2008 Jan;36(1):296-327.
(<http://www.survivingsepsis.org/implementation/resources/guidelines> 2008年8月4日アクセス)

[Shea 1996] Shea JA, Healey MJ, Berlin JA, Clarke JR, Malet PF, Staroscik RN, Schwartz JS, Williams SV. Mortality and complications associated with laparoscopic cholecystectomy. A meta-analysis. *Ann Surg*. 1996 Nov; 224(5):609-20.

[白岩ら 2009]白岩健、福田敬、渡辺茂、津谷喜一郎:イギリスNICEにおける医療技術評価の現状と医療技術ガイダンスのレビュー. *医療経済研究* Vol.21 (No.2). 155-170. 2009.

(<http://www.google.co.jp/url?sa=t&rct=j&q=Interventional+procedures+guidance&source=web&cd=7&ved=0CGoQFjAG&url=http%3A%2F%2Fwww.ihep.jp%2Fpublications%2Fstudy%2Fsearch.php%3Fd1%3D91%26i%3D6&ei=XkYbT4XeDMTwmAX6udWsCg&usg=AFQjCNGhi2hGy-hTu6aC6vflKqR0YuzSsA&sig2=S4rx1CHvdBAUeMtX-WIvAQ> 2012年1月22日アクセス)

[STA 2012]Guide to the single technology appraisal (STA) process. 2006.
(<http://www.nice.org.uk/niceMedia/pdf/tec>

hnologyappraisalprocessfinal.pdf 2012年1月22日アクセス)

[STROBE 2007] Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007 Nov;18(6):805-35. (観察的疫学研究報告の質改善(STROBE)のための声明:解説と詳細:
<http://www.strobe-statement.org/fileadmin/Strobe/uploads/translations/STROBE-Exp-JAPANESE.pdf> 2011/8/19アクセス)

[津田 2001] 津田敏秀, 馬場園明, 茂見潤, 大津忠弘, 三野善央. 医学における因果関係の推論—意思決定. *産業衛生学雑誌* 43(5), 161-173, 2001-09-20.
(<http://ci.nii.ac.jp/naid/1100038380632011/8/20>アクセス)

[津田 2009] 鈴木越治, 小松裕和, 頼藤貴志, 山本英二, 土居弘幸, 津田敏秀. 医学における因果推論 第一部—研究と実践での議論を明瞭にするための反事実モデル—. *日本衛生学雑誌* . Vol. 64 (2009), No. 4 p.786-795.
(http://www.jstage.jst.go.jp/article/jjh/64/4/64_786/_article/-char/ja 2011/8/20アクセス)

[USPSTF 2009] Petitti DB, Teutsch SM, Barton MB, Sawaya GF, Ockene JK, DeWitt

T; U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. *Ann Intern Med*. 2009 Feb 3; 150(3):199-205.

[湯浅 1998] 湯浅秀道, 浜島信之: 臨床研究における妥当性の概念. *最新医学*, 53: 483-491, 1998.

[Yuasa 2002] Yuasa H. The CONSORT statement: explanation and elaboration. *Consolidated Standards of Reporting Trials*. *Ann Intern Med*. 2002 Jun 18;136(12):926-7; discussion 926.

[湯浅 2004] 湯浅秀道. 「経験」が根拠となるには. *日本歯科評論* 64(4), 2004:52-58.

[湯浅 2011] 湯浅秀道. 「診療ガイドライン作成のルールの一つであるGRADEシステムについて: EBMの実践に役立つ資料とする視点より」. 平成22年度、厚生労働科学研究費補助金(地域医療基盤開発推進研究事業). 今後のEBM普及促進に向けた診療ガイドラインの役割と可能性に関する研究. 2011.

[Ward 2009] Ward AC. The role of causal criteria in causal inferences: Bradford Hill's "aspects of association". *EpidemiolPerspectInnov*. 2009 Jun 17;6:2.
(<http://www.epi-perspectives.com/content/6/1/2> 2011/8/19アクセス)