



## GRADE\*ガイドライン作成の流れの解説

\*GRADE: Grading of Recommendations Assessment, Development and Evaluation

本内容は、[GRADE-Jpn ML / ブログ](#)における解説内容(2012 年 6 月～7 月)であり、GRADE システムを利用した診療ガイドライン<sup>1</sup>の作成の流れを理解するために作成しました。各セクションの項目番号は以下の(1)ファイルに対応しており、「GRADE システムの使い方」\_20120426.ppt、および GRADE 教科書である「診療ガイドラインのための GRADE システム」(相原守夫 他、凸版メディア、2010 年)を利用して診療ガイドライン作成の流れを解説しています。

(1) 「コクランレビュー作成における GRADE システム(アプローチ)の利用・GRADE システムを利用した診療ガイドライン作成チェックリスト:治療介入編 Ver. 2 2012.7.17」

(pdf ファイル名: grade\_gl\_flow\_2012.pdf)

湯浅秀道先生(国立病院機構、豊橋医療センター、歯科口腔外科)と GRADE-Jpn ML メンバーが作成したチェックリストです。pdf 対応のエクセルファイルも公開しています。

[http://www.grade-jpn.com/grade\\_gl\\_flow\\_2012.pdf](http://www.grade-jpn.com/grade_gl_flow_2012.pdf)

[http://www.grade-jpn.com/grade\\_gl\\_flow\\_ma.xls](http://www.grade-jpn.com/grade_gl_flow_ma.xls)

(2) 「GRADE システムの使い方」(PPT ファイル名: how\_to\_use\_grade\_aihara\_20120426.ppt)

[http://homepage3.nifty.com/aihara/how\\_to\\_use\\_grade\\_aihara\\_20120426.lzh](http://homepage3.nifty.com/aihara/how_to_use_grade_aihara_20120426.lzh) より圧縮

版の DL 可能

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

GRADE システムに関して日本語で解説している唯一の書籍です。

(以下 GRADE 教科書と略)

相原守夫

平成 24 年 6 月 26 日～9 月 13 日

---

<sup>1</sup>米国医学研究所(Institute of Medicine:IOM)の 1990 年の定義によると、診療ガイドラインは「特定の臨床状況において適切な判断を行うため、臨床家と患者を支援する目的で、系統的に作成された文書」である。

2011 年 3 月、IOM は、「診療ガイドラインは患者ケアを最適化するための推奨を含む文書であり、エビデンスのシステマティック・レビューと、他の選択肢の利益と害の評価によって作成される」という新たな声明を発表した。

(“[Report: Clinical Practice Guideline We Can Trust](#)” )。

## 目次

はじめに:

- 1 [GRADE が登場した背景](#)
  - 2 [なぜ GRADE が必要なのか](#)
- Item #- 0: [ガイドラインパネル編成、プロセス確立](#)
- Item #- 1: [患者・臨床医の疑問\(CQ\)収集](#)
- Item #- 2: [疑問の定式化](#)
- Item #- 3: [アウトカムの選択](#)
- Item #- 4: [アウトカムの相対的な重要度の判定](#)
- Item #- 5: [疾患定義・選択基準など](#)
- Item #- 6: [SR:網羅的に検索する](#)
- Item #- 7: [SR:適格基準\(選択基準と除外基準\)](#)
- Item #- 8: [SR:アウトカムごとにデータを収集する](#)
- Item #- 9: [SR:個々の研究および研究全般のrisk of biasを評価する](#)
- Item #-10: [SR:コクランの色テーブル\(risk of summary\)を作成する、](#)  
附: [Risk of bias excel](#)
- Item #-11: [SR:アウトカムごとに risk of bias summaryを作成する](#)
- Item #-12: [SR:コクランの色グラフ\(横棒で%記載の方: risk of bias graph\):複数の研究全般のRoB](#)
- Item #-13: [SR:アウトカムごとに、9～12を繰り返し、複数の研究全般にわたるRoBを評価する](#)
- Item #-14: [メタ解析をする](#)  
附: [メタアナリシスのためのソフトウェア](#)
- Item #-15: [エビデンスの質:研究デザイン](#)  
附: [GRADEシステムによるエビデンスの質評価と定義](#)
- Item #-16: [エビデンスの質:\(グレードダウンの5要因\) limitations \(risk of bias\)](#)
- Item #-17: [エビデンスの質:\(グレードダウンの5要因\) inconsistency](#)
- Item #-18: [エビデンスの質:\(グレードダウンの5要因\) indirectness](#)
- Item #-19: [エビデンスの質:\(グレードダウンの5要因\) imprecision](#)
- Item #-20: [エビデンスの質:\(グレードダウンの5要因\) publication bias](#)
- Item #-21: [エビデンスの質:\(グレードアップの3要因\) 効果の程度が大きい](#)
- Item #-22: [エビデンスの質:\(グレードアップの3要因\) 交絡因子](#)
- Item #-23: [エビデンスの質:\(グレードアップの3要因\) 用量反応勾配](#)
- Item #-24: [各アウトカムについて、総体エビデンスの質\(body of evidence\)を判定する](#)
- Item #-25: [エビデンスプロファイル作成、SOF表作成](#)
- Item #-26: [医療資源に関する資料作成](#)
- Item #-27: [アウトカムの相対的な重要度の再評価](#)
- Item #-28: [そのCQの全体\(overall\)のエビデンスの質の判定](#)
- Item #-29: [推奨決定\(4要因\): 全体的なエビデンスの質](#)

Item #-30: [推奨決定\(4要因\)：望ましい効果と望ましくない効果のバランス](#)

Item #-31: [推奨決定\(4要因\)：価値観と好み](#)

Item #-32: [推奨決定\(4要因\)：資源の利用\(コスト\)](#)

Item #-33: [推奨度評価の合意形成\(RAND法；必要なら GRADEgrid\)](#)

Item #-34: [診療ガイドライン作成](#)

Item #-35: [レビューと外部評価](#)

Item #-36: [配布および普及](#)

Item #-37: [ガイドラインの管理](#)

Item #-38: [質の改善および実行](#)

最後に：

- [GRADEを採用する場合の注意](#)
- [GRADEを利用したとする基準](#)
- [GRADE JCE シリーズとは](#)

【はじめに】-(1) GRADEが登場した背景

ML発言番号00994

(2012/6/26)

Guyattらが、JAMA-Users' Guides 初版から述べていたにもかかわらず、エビデンスの質と推奨の評価において、「エビデンスの階層(……RCT 至上主義のような)」だけが注目されてしまって、診療ガイドラインが”研究デザインによって分類されたエビデンスの質”に依存するようになってしまいました。

この状況は、日本だけではなく海外でも共通したものでしたので、EBM working group の Guyatt と、当時 Cochrane のリーダーだった Oxman が一緒になって、2000年に GRADE working group を立ち上げたという歴史です。

Guyatt がたは、コクラングループが GRADE を採用し(2003年ごろ)、2008年の BMJ シリーズを発表して、これで適切な方向に向かうと思ったようですが、その期待は大きく裏切られ、今でも GRADE 改変論文や誤った GRADE 利用による診療ガイドラインが次々と発表されています。そこで、GRADE は、ガイドラインに特化した GRADE を解説すべく、JCE シリーズを 2011年から発表しているというわけです。2012年6月の時点では、20~22連載予定のうち、#12までが出版されています。

まとめると、

▷GRADE の登場してきた背景はなにか？

上記のような背景で、

”研究デザイン主体のエビデンス分類”と

”エビデンス依存の推奨決定”を修正することが目的です。

…「GRADE はエビデンスの質と推奨を分離する」

▷SR、GL とともに、GRADE はどのような状況に対して作られたのか？

”世界中にたくさんあるグレーディングシステムをひとつに統一する”

…「エビデンスの質評価と推奨決定プロセスを簡潔なものとして、その透明性を高める」

参考：

1. GRADE 教科書の FAQ

2. <http://www.grade-jpn.com/grade-faq-test1.html>

3. [http://www.grade-jpn.com/grade/guideline/grade-guideline\\_2008-nov.html#\[1\]](http://www.grade-jpn.com/grade/guideline/grade-guideline_2008-nov.html#[1])

[目次へ](#)

**【はじめに】-(2) なぜ GRADE が必要なのか**

ML 発言番号 00--- (6/27)

1. EBM を指導する方の視点
2. 患者の視点
3. 臨床医の視点
4. 政策決定者の視点

[1] EBM を指導する方の視点では;

医学文献からの論文を診療のガイドとして使用するための通常の3ステップアプローチ(結果は妥当か、結果は何か、患者ケアに適用できるか)のほとんどを、GRADE の枠組みで評価できる、ということが「エビデンスの質評価に GRADE を使う論拠」です。(スライド #11)

しかも、評価をひとつの表(エビデンスプロファイル: EP)に提示すると、長々とした文章を読まなくても一目瞭然で、研究の数、患者数、イベント数、そして効果の大きさやエビデンスの質評価の理由を理解できるというわけです。

EP の例; [http://homepage3.nifty.com/aihara/EP\\_heparin\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/EP_heparin_aihara_20120426.html)

効果推定値の確信の程度が、“エビデンスの質”です。

[2] 患者の視点では;

GRADE による推奨の強さは、strong か weak (conditional)の 2 種類のいずれかです。

“強い推奨”ならば、ほぼ全員が推奨される介入を希望するでしょう。“弱い推奨”ならば、多くの患者が提案される介入を希望するだろうし、希望しない患者も多い、ということです。

[3] 臨床医の視点では;

“強い推奨”ならば、ほぼ全ての患者が推奨される介入を受けるようにすべきであり、ガイドラインに準じた推奨を遵守しているかどうかは医療の質評価にも利用できます。“弱い推奨”ならば、推奨される介入を提案し、患者が意思決定できるように支援する必要があり、価値観や好みに基づいてエビデンスやエビデンスの要約を各自で吟味することになります。

[4] 政策決定者の視点では;

“強い推奨”ならば、ほとんどの状況下で、推奨事項をパフォーマンス指標として政策に採用できます。“弱い推奨”ならば、政策決定のためには、多数の利害関係者を含めてより実質的な論議を重ねる必要があります。

ちなみに、GRADE を採用している ACCP は、パフォーマンス指標は、GRADE の 1A、1B と評価された推奨事項を基準とすべきで、1C、2A、2B、2C と評価された推奨事項はパフォーマンス指標としては使用してはならないとしています(\*)。

参考:

1. GRADE 教科書 p98
2. ACCP ガイドラインのパフォーマンス指標と免責事項の記載 GRADE 教科書: p127、

■ Item #-0: ガイドラインパネル編成、プロセス確立

発言番号 00953

Blog: 2012/06/22

ガイドライン作成のためのグループ編成は(WHO に準じた)以下の仮想チームとしています。  
(GRADE 教科書の p148 参照)

- (1)ガイドライン運営委員会
- (2)専門ガイドライン作成グループ(パネル)
- (3)特別作業班(タスクフォース)
- (4)事務局
- (5)外部専門家

当仮想パネルには、癌専門医、臨床医、統計学者、ガイドライン作成経験のある人、システマティック・レビュー作成者(最低 2 人)費用効果分析者、エンドユーザーである患者(あるいは患者代表者)、そして GRADE working group member がそろっています。……すばらしいチームです！

=====

注:

- ・ガイドラインパネルは全員、利益相反の公表が必要(GRADE 教科書 p152)。
- ・パネル全員は、GRADE を使うということに同意する必要があります。  
これはとても重要なことであり、ある介入に関する推奨を決定する際の投票などに参加すべきではないメンバーがでてくる可能性があります。  
例えば、ある薬剤の開発に関与した人は、エビデンスの質評価には参加できても、当該薬剤の推奨決定には関与できない、ガイドラインに示す推奨決定結果によってメンバー脱退ということは許されないという原則です。当然ながら、ガイドライン作成者リストに名前だけを乗せるだけの参加はよろしくないということです。

参考:

1. ガイドラインパネル編成、プロセス確立  
[http://www.grade-jpn.com/grade/guideline/grade-guideline\\_2008-nov.html#\[1\]](http://www.grade-jpn.com/grade/guideline/grade-guideline_2008-nov.html#[1])
2. 利益相反の公表 GRADE 教科書 p152

これらは、上記リンク「#1」の下図に示す ref #3 (GRADE システムがはじめて、公開された WHO 資料)を翻訳したものです。

[1]: GRADEとは: ガイドラインパネル編成、プロセス確立: [\[top\]](#) [\[FAQ翻訳へ\]](#)  
[\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#)

- ・ **GL作成グループ**は、いくつかのチームから構成されているべきである。(WHO-GWG): Guideline Steering Group Secretariatなど。
- ・ ガイドライン担当者は、利益相反(conflict of interest)を明示すべきである。なんらかの利益相反という理由でが考。Annex-A
- ・ [Ref]:
  - 1.Cochrane Handbook (ver5).
  - 2.Improving the use of research evidence in guideline development (SERIES) (no.1-16)(2006)
  - 3.GWG (guidelines for WHO guidelines, 2003)

■ Item #-1: 患者・臨床医の疑問(CQ)収集

Blog: 2012/07/02

GRADE ガイドラインの作成の流れを説明するために利用するシナリオは「GRADE システムの使い方」\_20120426.ppt にあるように、(2012/6/22 に内容を改変した how\_to\_use\_grade\_aihara\_20120426.lzh を web 上にアップしましたので、古いデータ利用の方は気をつけてください)、

スライド #17 にあるようなシナリオから、1つの臨床疑問を作成して GRADE ガイドラインを作成してみるという作業です。

“固形癌患者において、非経口的抗凝固療法を実施すると癌患者の死亡率が低下できて生存延長につながるのだろうか？”という“治療”に関する疑問です。

癌患者における非経口的抗凝固療法(ヘパリン):  
生存延長のために使用すべきか

腫瘍内科医であるあなたは、外来で肺癌や胃癌、大腸癌、前立腺癌の患者を診察している。  
静脈血栓塞栓症を合併していないこれらの固形癌患者に対して、非経口的なヘパリン(未分画ヘパリン、低分子ヘパリンなど)使用は、癌患者の生存を延長させるだろうか。  
ヘパリン投与による深刻な出血はないだろうか。  
QOLや血小板減少はどうだろうか。

臨床研究には、目的に応じて5つの分野(治療、害、鑑別診断、診断、予後)があり、それぞれに適した研究デザインがある。たとえば、治療に関する研究にはランダム化比較試験(RCT)やシステマティック・レビュー、害の研究にはRCTや観察研究、診断研究には横断研究や RCT などが適している。(GRADE教科書 p29)

表 1.1-1 ヘルスケア・クエスチョンの種類

疑問の種類	意味	入手可能なエビデンスの最良の研究デザイン
治療	“患者にとって重要なアウトカム”(症候、機能、生存率、死亡率、コスト)についての介入効果を決定する。	RCT, システマティック・レビュー
害	有害の可能性がある曝露(治療を含む)の影響を確認する。	RCT, 観察研究(コホート研究, 症例対照研究)
鑑別診断	特有の臨床徴候をもつ患者において、基礎疾患の頻度を確認する。	観察研究
診断	着目している状況や疾患のある者とない者を区別する検出力を確認する。	RCT, 横断研究
予後	患者の病気が今後どうなっていくか予測する。	コホート研究

注: クリニカル・クエスチョン(臨床上の疑問、臨床疑問)という用語が EBM では多用されますが、GRADE 論文においては、より広い意味のヘルスケア・クエスチョンが記載されることもあります。これは、“対象が必ずしも患者ではないことがある”という理由です。(GRADE 教科書 p28)

■ Item #-2: 疑問の定式化

Blog: 2012/07/03

ほとんどの疑問は、患者(P)、介入(I)、比較(C)、アウトカム(O)に分けられる。

患者の治療について決断をしなくてはならない臨床医が想起するヘルスケア・クエスチョン(特に前景疑問)は、関連する患者集団、治療方法や曝露、および、治療や曝露によって生じる“患者にとって重要なアウトカム”を特定したものであり、基本的にはPICOの4要素から成っている。

*非経口的抗凝固療法は癌患者の生存延長のために使用すべきか*

**Patient:** 癌患者(肺癌、大腸癌、前立腺癌など)

**Intervention:** parenteral anticoagulation (ヘパリン)

**Comparison:** プラセボ、no heparin

**Outcomes:** 全死亡率、大出血、小出血、症候性 DVT、QOL、血小板減少

(注)

通常は、PICO 形式の中の特にPIO (Patient、Intervention、Outcomes)について明確にして検索すると、関連するエビデンスを見つけやすい。

### ■ Item #3: アウトカムの選択

Blog: 2012/07/05

アウトカムとは転帰や帰結のことで、研究の目的測定項目を意味する。特に臨床研究では評価項目・指標を意味するエンドポイントという用語を用いて、主要(1次)エンドポイント、副次(2次)エンドポイント、複数の指標を組み合わせた複合エンドポイントなどの表現が使われる。アウトカム変数には、2値(dichotomous)変数と連続(continuous)変数の2種類がある。

患者にとって重要なアウトカムとは、次のような質問に「はい」という答えが出るようなアウトカムと定義される。

「患者が、この治療によって変化する唯一のアウトカムがこのアウトカムであると知った場合、それに副作用やコストを伴うのだとしても患者はその治療を受けることを考慮するだろうか。」

#### スライド #20:

患者にとって重要なアウトカムとは、以下のカテゴリーの I、II、III であり、IVは患者にとって重要なものではない。

<b>I. Mortality</b>
1. all cause mortality
2. disease specific mortality
<b>II. Morbidity</b>
1. cardiovascular major morbid events
2. other major morbid events (e.g. loss of vision, seizures, fracture, revascularization)
3. onset/recurrence/relapse/remission of cancer and other chronic diseases (e.g. COPD exacerbation, new onset of diabetes)
4. renal failure requiring dialysis
5. hospitalization, medical and surgical procedures (e.g. placement of a pacemaker, and cardioversion)
6. infections
7. dermatological/ rheumatologic disorders
<b>III. Symptoms/Quality of life/Functional status</b> (e.g. failure to become pregnant, successful nursing/breastfeeding, depression)
<b>IV. Surrogate outcomes</b> (e.g. diagnosis of TB, viral load, physical activity, weight loss, post operative atrial fibrillation, cognitive function, recurrent (malignant) polyps, adherence to medication)

本シナリオでは、(システマティック・レビューにおける「重要」なアウトカム)

Outcomes: 全死亡率、大出血、小出血、症候性 DVT、QOL、血小板減少

(注)

患者にとって重要なアウトカムに関する情報が入手できない状況では、代替アウトカムを考慮すべきであるが、ヘルスケア・クエスチョンのリストには代替アウトカムではなく本来の重要なアウトカムを記載すべきである。

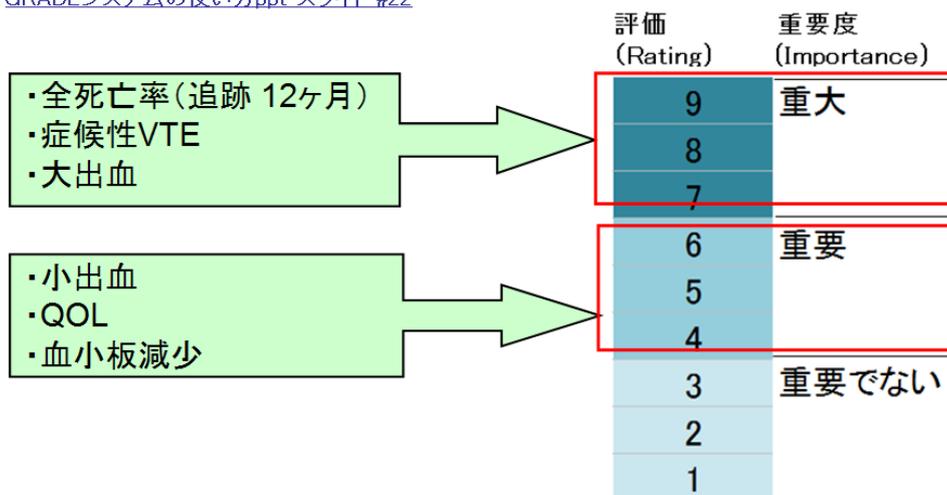
■ Item #-4: アウトカムの相対的な重要度の判定

Blog: 2012/07/05

GRADE では、アウトカムの重要度を、9段階の点数で相対的に評価し、それらを3点きざみで3つに分類する。意志決定にとって、「重大:7～9点」、「重要:4～6点」、「重要ではない:1～3点」の分類である。

この3つの分類の中でシステマティック・レビューに関係するアウトカムは、「重大」なものや「重要」なものであり、両者をGRADEエビデンステーブルに含まなければならない。ガイドラインの推奨に影響するアウトカムは、「重大」なアウトカムのみである。

GRADEシステムの使い方ppt スライド #22



(注)

1.通常、どのアウトカムが重要なのかについては、システマティック・レビューやガイドライン作成に着手する前の、研究計画書作成の段階で決定する必要がある。その一方で、ガイドラインパネルは、研究計画書の作成後または分析の実施後になってアウトカムの重要度を再評価すべきである。(GRADE 教科書 p32 参照)

2.診療ガイドライン作成においては、「重大」なアウトカムだけを検討するため、小出血や QOL、血小板減少のアウトカムは考慮しない。

■ Item #-5: 疾患定義・選択基準など

Blog: 2012/07/06

疾患や介入の定義とともに、採用するエビデンスの種類 (研究デザイン) を決断する。異なるタイプの疑問に取り組むためにはどんな種類のエビデンスが使われるべきかを明確にすることが大切である。

- **The study designs** to be included in a review should be dictated by the interventions and outcomes being considered. A decision about how broad a range of study designs to consider should be made in relationship to the characteristics of the interventions being considered, what evidence is available, and the time and resources available.
- Decisions about the range of study designs to include should be made explicitly.
- Great caution should be taken to avoid confusing a lack of evidence with evidence of no effect, and to acknowledge uncertainty.
- **Expert opinion is not a type of study design and should not be used as evidence.**

引用文献:

Improving the use of research evidence in guideline development: 7. Deciding what evidence to include Oxman AD, Schünemann HJ and Fretheim A

<http://www.health-policy-systems.com/content/pdf/1478-4505-4-19.pdf>

シナリオの場合、以下のようになる。

**Types of studies:** Randomized controlled trials (RCTs)

**Types of participants:**

Patients with cancer with no indication for prophylactic or therapeutic anticoagulation

**Types of interventions:**

Experimental intervention: parenteral anticoagulants such as UFH, LMWH and fondaparinux

Comparator intervention: placebo or no intervention.

**Outcomes:**

All-cause mortality, Symptomatic venous thromboembolism, Health-related quality of life、

Major bleeding, Minor bleeding, Thrombocytopenia

■ Item #-6: SR: 網羅的に検索する

Blog: 2012/07/10

システマティック・レビューとは、「明確に定式化された疑問について、関連する研究の特定・選択・批判的吟味、および採用研究からのデータを集めて解析する、系統的で明確な方法を用いるレビュー」である。

つまり、システマティック・レビューという言葉は、バイアスの可能性を減少するようにデザインした方法を用いて、臨床疑問、特に前景疑問を扱う要約ということである。非系統的なアプローチでは、誤り（ランダム誤差）だけではなくバイアス（系統誤差）も生じやすく、介入効果などが過小評価されたり過大評価されたりする。

関連する研究の特定・選択では、コクランライブラリのような既存のシステマティック・レビューや、PubMed、Ovid などの電子検索、引用検索、ハンドサーチなど徹底的な検索を行って、さらに必要に応じて専門家に連絡を取るなど、文献入手のためにバイアスやエラーを少なくする努力が必要である。また、少なくとも1つのデータベースの電子検索式について、使用されたあらゆる“limits”を含め、再現できるくらいに詳細に示すことが大切である (PRISMA)。

参考:

1. エビデンスの検索とデータの抽出・統合: GRADE教科書 p36
2. GRADE ガイドラインにおけるシステマティック・レビュー

[http://www.grade-jpn.com/grade/guideline/grade-guideline\\_2008-nov.html#\[4\]](http://www.grade-jpn.com/grade/guideline/grade-guideline_2008-nov.html#[4])

シナリオにおける検索式 (スライド #25)

MEDLINE

```
#1 Heparin/  
#2 Heparin.tw  
#3 Heparin, Low-Molecular-Weight/  
#4 (LMWH OR low molecular weight heparin OR nadroparin OR  
fraxiparin OR enoxaparin OR clexane OR lovenox OR dalteparin OR  
fragmin OR ardeparin OR normiflo OR tinzaparin OR logiparin OR  
innohep OR certoparin OR sandoparin OR reviparin OR divarin OR  
danaproid OR orgaran).tw  
#5 1 OR 2 OR 3 OR 4  
#6 Coumarins/  
#7 Warfarin/  
#8 (warfarin OR coumadin OR acenocumarol OR phenprocumon OR 4-  
hydroxicoumarins OR oral nticoagulant OR vitamin K antagonist OR  
VKA).tw  
#9 6 OR 7 OR 8  
#10 (fondaparinux OR Arixtra).tw  
#11 (ximelagatran OR Exanta).tw  
#12 (Pradaxa or Dabigatran or rivaroxaban or Xarelto or apixaban).tw.  
#13 5 OR 9 OR 10 OR 11 OR 12  
#14 Neoplasms/  
#15 (malignan$ OR neoplasm$ OR cancer OR carcinoma$ OR  
#16 14 OR 15  
#17 clinical trial.pt. OR random:.tw. OR tu.xs.  
#18 animals/ NOT human/  
#19 17 NOT 18  
#20 13 AND 16 AND 19
```

## ■ Item #-7: SR: 適格基準(選択基準と除外基準)

Blog: 2012/07/10

適格基準はデータ抽出前に作成する。適切な適格基準はバイアスを排除し、メタアナリシスの再現性を保障するものである。適格基準の項目は、メタアナリシスのテーマや目的によってさまざまであるが、基本的な項目は共通している。

適格基準の項目	説明
1.研究デザイン	基本的には、同一の研究を統合する。RCTと非ランダム化試験は統合すべきでない。
2.出版・研究の時期	選択した理由とともにその時期を明記する。新しい情報をなるべく含むようにする。
3.言語	実際は、英語が対象になるケースが多くなる。英語抄録のない他言語の文献情報を得るためには、論文を入手し翻訳する必要がある、費用も時間もかかる。(*)
4.複数発表における選択基準	同じ被験者のデータが2つ以上の論文に掲載されることがある。これらの論文データをメタ解析で統合するのは誤りである。複数発表の研究資料がある場合は、その中からどれを選択するかは必要である。
5.サンプル数・観察期間	統合結果に大きな影響を及ぼす傾向がある方法(例、general variance-based method)を用いる場合は、サンプル数の少ない研究を除外する基準を作成する。サンプル数、観察期間については、適格基準による不適格論文の除外を行わず、統計処理ののちに検討する感度分析を実施してもよい。
6.類似治療・介入および類似効果の許容範囲	治療や介入およびその効果を統合する際に、どの程度まで同じとみなすのかは重要なポイントである。条件を厳しくすると、適格基準を満たす論文が少なくなり、「検出力」が低下する。
7. 1～6の基準を満たした不完全な報告の適格性	上記の適格基準を全て満たしているものの、学会発表抄録やレターなど、研究論文と比較して不完全な報告がある。これらは、適格基準を満たすかどうか判断できないこと、査読が行われていない可能性があることから、不適格になりえる。しかし、公開バイアスを考えると含めるほうが適切かもしれない。感度分析を行うという方法もある。

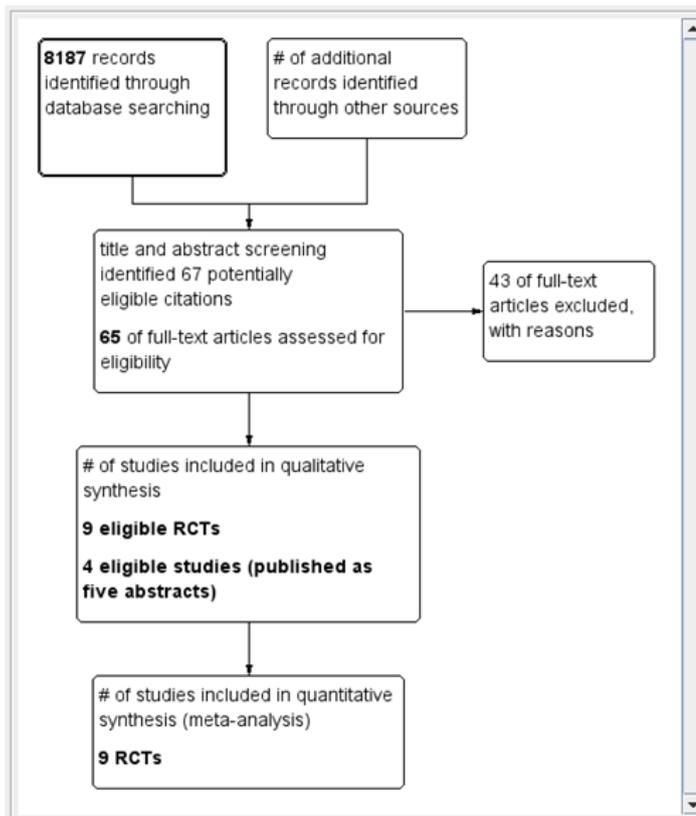
(\*) 英語で記述された報告と英語以外の言語で報告されたものを比較すると、言語制限のないメタ解析のほうがより精確という報告や、特定の国からの RCT 報告は、ほとんどが統計的に有意な結果のものだけであると示唆するエビデンスがある。

シナリオの場合、

検索した 8, 187 件の論文から、最終的に 9 件の RCT を選択した。(PRISMA フローチャート)

PRISMA フローチャート (スライド #26)

PRISMA flowchart



検索した8187件の論文から、最終的に 9件のRCTをメタアナリシスのために選択した。

1	<a href="#">Agnelli 2009</a>
2	<a href="#">Altinbas 2004</a>
3	<a href="#">Kakkar 2004</a>
4	<a href="#">Klerk 2005</a>
5	<a href="#">Lebeau 1994</a>
6	<a href="#">Pelzer 2009</a>
7	<a href="#">Perry 2010</a>
8	<a href="#">Sideras 2006</a>
9	<a href="#">Weber 2008</a>

■ Item #-8: SR:アウトカムごとにデータを収集する

Blog: 2012/07/12

システマティック・レビューにおけるデータの収集(参加者、介入、比較介入、研究デザインなど)はアウトカムごとに実施する。

個別の研究単位でエビデンスの質の評価を行うという誤解が国内にはあるが、GRADEのエビデンス評価は原則として複数の研究からのシステマティック・レビューをもとに実施するものであり、個別の研究単位の評価は行わない。PRISMA チェックリストにおいて、「アウトカムレベル」の評価という新たな概念が導入されている。(GRADE 教科書 p176)

エビデンスプロファイルや SoF に含むべき意思決定に重要なアウトカムの個数は7個を上限とする(コクランハンドブック)。検索して論文がないアウトカムも、重要・重大ならば、エビデンスプロファイルに記載すべきである。

シナリオの場合、(スライド #27)

アウトカム-1 (all-cause mortality) に関しては 9 件、アウトカム-4 (症候性 VTE) に関しては 7 件、アウトカム-5 (QOL) に関しては1件がレビューに組み込まれている。

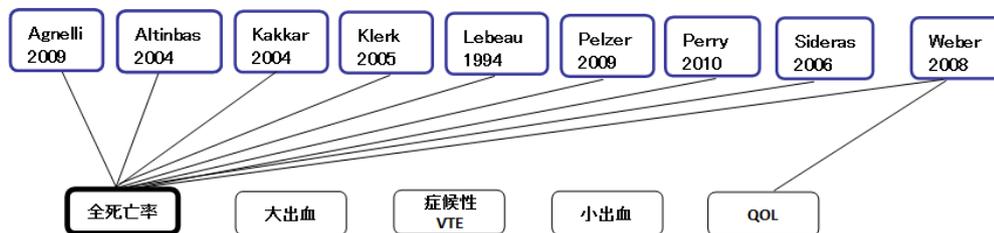
Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation (Review)  
 Akl EA, Gunukula S, Barba M, Yosuiico VED, van Doormaal FF, Kuipers S, Middeldorp S, Dickinson HO, Bryant A, Schünemann H  
<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006652.pub3/pdf>

	研究	アウトカム 1	アウトカム 2	アウトカム 3	アウトカム 4	アウトカム 5
1	<a href="#">Agnelli 2009</a>	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
2	<a href="#">Altinbas 2004</a>	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
3	<a href="#">Kakkar 2004</a>	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
4	<a href="#">Klerk 2005</a>	all-cause mortality	major bleeding	minor bleeding	-	-
5	<a href="#">Lebeau 1994</a>	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
6	<a href="#">Pelzer 2009</a>	all-cause mortality	major bleeding	-	symptomatic VTE	-
7	<a href="#">Perry 2010</a>	all-cause mortality	major bleeding	-	symptomatic VTE	-
8	<a href="#">Sideras 2006</a>	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	QOL
9	<a href="#">Weber 2008</a>	all-cause mortality	major bleeding	minor bleeding	-	-

研究によっては、アウトカムの報告がないものがある

(-: 報告なし)

スライド #28



全死亡率のアウトカムは9件研究で報告されており、QOLのアウトカムに関しては1件の研究のみである。このように、アウトカムによって、統合する研究の数は異なることが通常である。

注：  
本図においては、大出血、症候性VTE、小出血に関しては表示していない。



あるアウトカムに関して、複数の研究を統合したメタアナリシスのエビデンスを、**body of evidence**と呼び、このエビデンスの質を評価する。

[目次へ](#)

■ Item #-9: SR:アウトカムごとに個々の研究および研究全般の risk of bias を評価する

Blog: 2012/07/13

バイアスとは、結論や推論における系統誤差、または真実からの逸脱である。バイアスには方向性があり、真の介入効果を過小評価させる方向に働くバイアスもあれば、過大評価させる方向に働くバイアスもある。通常、ある研究の結果がバイアスの影響を潜在的にどの程度受けているのかを把握することは困難である。そのため、バイアスのリスク(risk of bias)として検討したことを適切に記載し、考慮したことを明示することが大切である。

コクランレビューの「risk of bias」は、GRADEの「研究の限界」に相当するもので、以下のような基本的なドメインがある。

割付の順序 (random sequence generation)
割付の隠蔽 (allocation concealment)
参加者と研究者の盲検化 (blinding of participants and personnel )
アウトカム評価者の盲検化 (blinding of data analyst)
不完全なアウトカム (incomplete outcome data)
選択的な報告 (selective reporting)
その他のバイアス(stopped early for benefit, etc)

注:

sequence generation, allocation concealment、selective outcome reportingのドメインについては、RevMan旧版では、参加者、研究者、アウトカム評価者の盲検化に関わる複数のバイアスを全て1つの領域で評価していた(ただし、異なるアウトカムについては、アウトカム別にバイアスを評価していたかもしれない)。改訂版ツール(RevMan 5.1)では、参加者、研究者、アウトカム評価者の盲検化にかかわるバイアスは、アウトカム評価者の盲検化にかかわるバイアスとは別に評価するようになっている([コクランハンドブック 8.5.1 overviewの訳文も参照のこと](#))。

各研究のrisk of bias評価結果は「低 low」、「不明 unclear」、「高 high」の3つに分類される。risk of bias summaryやrisk of biasグラフでは、Low は緑で、Unclear は黄色、High は赤で表示される。



Figure 1. 'Risk of bias' summary - mortality at 12 months: review authors' judgements about each methodological quality item for each included study that evaluated mortality

	Adequate sequence generation?	Allocation concealment?	Blinding of patients?	Blinding of providers?	Blinding of data collectors?	Blinding of outcome adjudicators?	Blinding of data analysts?	Incomplete outcome data addressed?	Free of selective reporting?	Free of other bias?	Intention to treat analysis?
Agnelli 2009	+	+	+	+	+	+	-	+	+	+	+
Altinbas 2004	+	-	-	-	+	+	-	+	+	+	+
Kakkar 2004	+	+	+	+	+	+	-	+	+	+	+
Klerk 2005	+	+	+	+	+	+	-	+	+	+	+
Lebeau 1994	+	+	-	-	+	+	-	+	+	+	+
Pelzer 2009	+	+	-	-	+	+	+	+	+	+	+
Perry 2010	+	+	+	+	+	+	-	+	+	+	+
Sideras 2006	+	+	-	-	+	+	-	+	+	+	?
Weber 2008	+	+	-	-	+	+	-	+	+	+	+

\*参考：[コクランハンドブック 8.5.1 overview](#)

「順番の生成」、「割付けの隠蔽」、「選択的アウトカム報告」のドメインについては、いずれもこのRoBツールでそれぞれの研究に1つのエントリとして使用することにより対応すべきである。「参加者および研究者の盲検化」、「アウトカム評価の盲検化」、ならびに「不完全なアウトカムデータ」については、異なるアウトカム（あるいは異なる時点における同一のアウトカム）は個別の評価を必要とすることから、複数のエントリを使用する必要がある。レビュー著者はアウトカムをグループ化することによりエントリ数を制限する努力をすべきである。たとえば、「アウトカム評価の盲検化」を評価する場合は“主観的”または“客観的”アウトカムにグループ化する、あるいは「不完全なアウトカムデータ」の場合は“6ヶ月目の患者報告アウトカム”または“12ヶ月目の患者報告アウトカム”にグループ化する。これと同じアウトカムのグループ化を、レビューに含まれる各研究に適用する。最後のドメイン（「その他のバイアス」）については、複数の研究全体に関する1つのエントリとして評価するとよい（RevManの初期状態）。ただし、特定の「その他のrisk of bias」を評価する際にはあらかじめ指定されたエントリを使用することが強く推奨される。このような著者指定のエントリは、研究全体に対するエントリの場合、または各研究内の個々の（またはグループ化された）アウトカムに対するエントリの場合がある。

## ■ Item #-10: SR:コクランの色テーブル(risk of bias summary)を作成する

Blog: 2012/07/17

Risk of Bias テーブルは、レビューによる内的妥当性の評価（あるいは、研究においてリサーチクエスチョンに対する回答が適格に、またはバイアスなく提供されているかの評価）を支援するためのもので、ツールを使って作成される。ツールに入力される情報、および Risk of Bias テーブルに示される情報は、SoF テーブルにおいて各アウトカムのエビデンスの質を検討する際に重要となってくる。GRADEpro において GRADE を使用してエビデンスの質を評価する場合、著者はエビデンスの質を左右する 5 つの要因について検討する。そのうちの 1 つの要因としてあげられるのが、「デザインの限界」である。「デザインの限界」での妥当性への脅威は、Risk of Bias ツールで評価されるものと同様である。そのため、レビューは Risk of Bias テーブルの作成時に下した判断に基づき、GRADEpro における「デザインの限界」を理由にある特定のアウトカムのエビデンスの質の評価を下げるべきかどうかについて判断する必要がある。[GRADEpro の FAQ から](#)

個々の研究の RoB 評価を項目別に図示したものを、Risk of bias summary と呼び、縦横の表示を変えて、RoB 項目別に複数の研究の評価を%で図示したものを、Risk of bias graph と呼ぶ。

RevMan を使うことで、各研究の RoB table データから、アウトカム別の risk of bias summary と risk of bias graph が作成できる(アウトカムごとに RoB の入力を再度行い、作りなおす必要がある)。

例： 非経口的抗凝固療法は癌患者の生存延長のために使用すべきか

アウトカム-1(全死亡率)	研究1・2・3・4・5・6・7・8・9
アウトカム-2(大出血)	研究1・2・3・4・5・6・7・8・9
アウトカム-3(小出血)	研究1・2・3・4・5・8・9
アウトカム-4(症候性VTE)	研究1・2・3・5・6・7・8
アウトカム-5(QOL)	研究9



### risk\_of\_bias\_excel

あるアウトカムについて、個々の研究の Risk of bias を評価したテーブルから risk of bias summary や risk of bias graph を表示する流れは、エクセルをつかうと簡単に理解できる。

(参照 [risk\\_of\\_bias\\_excel.xls](#))

例；

アウトカム-1 (全死亡率) 9件の RCT		アウトカム-5 (QOL) 1件の RCT																																																																																	
	<table border="1"> <thead> <tr> <th></th> <th>割付の順序 (random sequence generation)</th> <th>割付の隠蔽 (allocation concealment)</th> <th>参加者と研究者の盲検化 (blinding of participants and researchers)</th> <th>アウトカム評価者の盲検化 (blinding of outcome assessors)</th> <th>不完全なアウトカム (incomplete outcome data)</th> <th>選択的な報告 (selective reporting)</th> <th>その他のバイアス (stopped early for other reasons)</th> </tr> </thead> <tbody> <tr><td>Agnelli 2009</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Altinbas 2004</td><td>Green</td><td>Red</td><td>Red</td><td>Red</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Kakkar 2004</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Klerk 2005</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Lebeau 1994</td><td>Green</td><td>Green</td><td>Red</td><td>Red</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Pelzer 2009</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Perry 2010</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td><td>Green</td></tr> <tr><td>Sideras 2006</td><td>Green</td><td>Green</td><td>Red</td><td>Red</td><td>Green</td><td>Green</td><td>Yellow</td></tr> <tr><td>Weber 2008</td><td>Green</td><td>Green</td><td>Red</td><td>Red</td><td>Green</td><td>Green</td><td>Green</td></tr> </tbody> </table>		割付の順序 (random sequence generation)	割付の隠蔽 (allocation concealment)	参加者と研究者の盲検化 (blinding of participants and researchers)	アウトカム評価者の盲検化 (blinding of outcome assessors)	不完全なアウトカム (incomplete outcome data)	選択的な報告 (selective reporting)	その他のバイアス (stopped early for other reasons)	Agnelli 2009	Green	Green	Green	Green	Green	Green	Green	Altinbas 2004	Green	Red	Red	Red	Green	Green	Green	Kakkar 2004	Green	Klerk 2005	Green	Lebeau 1994	Green	Green	Red	Red	Green	Green	Green	Pelzer 2009	Green	Perry 2010	Green	Sideras 2006	Green	Green	Red	Red	Green	Green	Yellow	Weber 2008	Green	Green	Red	Red	Green	Green	Green																										
	割付の順序 (random sequence generation)	割付の隠蔽 (allocation concealment)	参加者と研究者の盲検化 (blinding of participants and researchers)	アウトカム評価者の盲検化 (blinding of outcome assessors)	不完全なアウトカム (incomplete outcome data)	選択的な報告 (selective reporting)	その他のバイアス (stopped early for other reasons)																																																																												
Agnelli 2009	Green	Green	Green	Green	Green	Green	Green																																																																												
Altinbas 2004	Green	Red	Red	Red	Green	Green	Green																																																																												
Kakkar 2004	Green	Green	Green	Green	Green	Green	Green																																																																												
Klerk 2005	Green	Green	Green	Green	Green	Green	Green																																																																												
Lebeau 1994	Green	Green	Red	Red	Green	Green	Green																																																																												
Pelzer 2009	Green	Green	Green	Green	Green	Green	Green																																																																												
Perry 2010	Green	Green	Green	Green	Green	Green	Green																																																																												
Sideras 2006	Green	Green	Red	Red	Green	Green	Yellow																																																																												
Weber 2008	Green	Green	Red	Red	Green	Green	Green																																																																												
		Weber 2008	<table border="1"> <thead> <tr> <th></th> <th>割付の順序 (random sequence generation)</th> <th>割付の隠蔽 (allocation concealment)</th> <th>参加者と研究者の盲検化 (blinding of participants and researchers)</th> <th>アウトカム評価者の盲検化 (blinding of outcome assessors)</th> <th>不完全なアウトカム (incomplete outcome data)</th> <th>選択的な報告 (selective reporting)</th> <th>その他のバイアス (stopped early for other reasons)</th> </tr> </thead> <tbody> <tr><td>Weber 2008</td><td>Green</td><td>Green</td><td>Red</td><td>Red</td><td>Green</td><td>Green</td><td>Green</td></tr> </tbody> </table>		割付の順序 (random sequence generation)	割付の隠蔽 (allocation concealment)	参加者と研究者の盲検化 (blinding of participants and researchers)	アウトカム評価者の盲検化 (blinding of outcome assessors)	不完全なアウトカム (incomplete outcome data)	選択的な報告 (selective reporting)	その他のバイアス (stopped early for other reasons)	Weber 2008	Green	Green	Red	Red	Green	Green	Green																																																																
	割付の順序 (random sequence generation)	割付の隠蔽 (allocation concealment)	参加者と研究者の盲検化 (blinding of participants and researchers)	アウトカム評価者の盲検化 (blinding of outcome assessors)	不完全なアウトカム (incomplete outcome data)	選択的な報告 (selective reporting)	その他のバイアス (stopped early for other reasons)																																																																												
Weber 2008	Green	Green	Red	Red	Green	Green	Green																																																																												

(本解説においては、Akl EA 論文の RoB 評価項目を改変した)

参考：

1. 研究全体でのコクラン risk of bias の考え方: GRADE 教科書 p50
2. スライド #39～#45

■ Item #-11: SR:一つの研究の (within study)、そのアウトカムに関する RoB の確認

Blog: 2012/07/18

主要基準の何%がリスクが高いと判定されたらという画一的なものでない。

“論文そのもの” に対して、「この論文は、low risk of bias」などという評価するのでなく、あくまでも、「この論文は、” このアウトカムについて” low risk of bias 研究といえる」と評価するだけである。

例： 単一研究の RoB summary (アウトカム:死亡率)

[risk of bias excel.xls](#) を使用

アウトカム	死亡率		
Sideras 2006	bias	判定	コメント
	割付の順序 (random sequence generation)	low	
	割付の隠蔽 (allocation concealment)	low	
	参加者と研究者の盲検化 (blinding of participants and personnel)	high	
	アウトカム評価者の盲検化 (blinding of data analyst)	high	
	不完全なアウトカム (incomplete outcome data addressed)	low	
	選択的な報告 (selective reporting)	low	
	その他のバイアス (stopped early for benefit, ITT)	unclear	

Risk of bias summary	
割付の順序 (selection bias)	割付の隠蔽 (selection bias)
参加者と研究者の盲検化 (performance bias)	アウトカム評価者の盲検化 (detection bias)
不完全なアウトカム (attrition bias)	選択的な報告 (reporting bias)
その他のバイアス (利益のため早期中止など)	
Sideras 2006	low, high, low, low, unclear

(本解説においては、Akl EA 論文の評価項目を改変した)

[目次へ](#)

■ Item #-12: SR:コクランの色グラフ(横棒で%記載の方: risk of bias graph)を作成する:  
そのアウトカムに関する研究全般の RoB

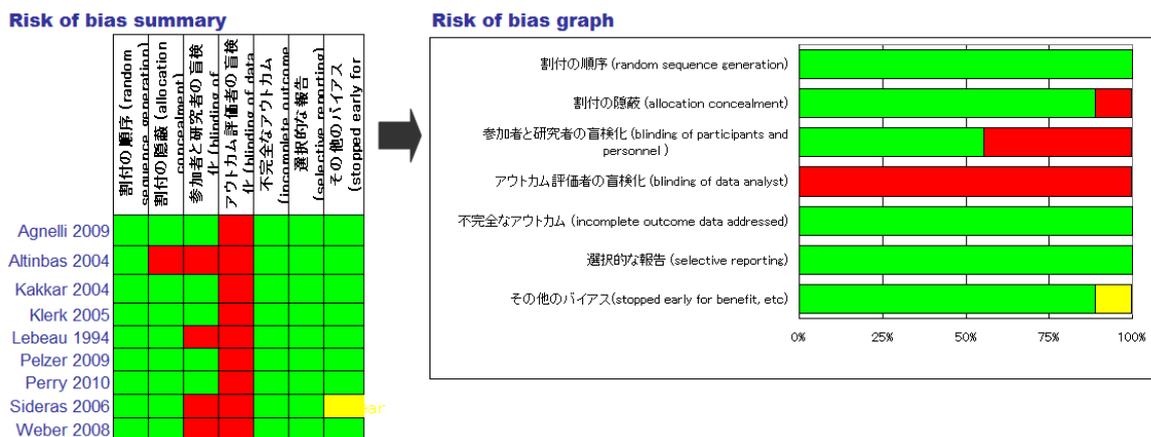
Blog: 2012/07/18

risk of bias graph は risk of bias summary 図の縦横の表示を変えただけである。  
すなわち、評価項目別に全研究の RoB を集計したものが、risk of bias graph である。

アウトカムごとに個別研究の RoB を評価し、その後複数の研究にわたる RoB を判定する。  
この複数の研究にわたる RoB が、GRADE の"limitations"(あるいは、risk of bias)と同じものであり、つまり、グレードダウンの 5 要因の 1 つである。

Review Manager (RevMan) のソフトを利用する事が望まれるが、エクセルを使っても解析可能である(下図では、[risk of bias excel.xls](#) 使用)

シナリオ: アウトカム(全死亡率)についての risk of bias graph



(本解説においては、Akl EA 論文の評価項目を改変した。スライド #44 参照)

“割付の隠蔽“の評価は、研究全体で見ると、89%(8/9)が緑色の low risk of bias である。

“アウトカム評価者の盲検化“は、全ての研究で(9/9)、赤色の high risk of bias である。

注:

全死亡の評価は生死という明確なアウトカム評価であることから、評価者の盲検化はそれほど重要なことではない(エビデンスの質を下げることはない)

■ Item #13: SR:アウトカムごとに、9～12を繰り返し、複数の研究全般にわたる RoB を評価する

Blog: 2012/07/18

アウトカムごとに複数の研究にわたって (across studies)、RoB を評価する。

アウトカムによって各研究の RoB 評価が異なることがある。

主要基準の何%がリスクが高いと判定されたらという画一的なものでない。

評価の重要ポイントは、脚注に記載すること。

GRADE 教科書 p51 に、RCT における risk of bias の評価基準を記載してあるが、

GRADE working group は複数の研究にわたる RoB 評価のために、以下のような指針を発表している(スライド #42)。

個々の研究のコクランrisk of biasテーブルに基づき、複数の研究にわたる全体的なエビデンスの質 (overall quality of evidence across studies) を評価を行うための指針 (GRADE working group)

第一に、全体的なエビデンスの質を決定するにあたり、複数の研究の平均値をとってはならない(たとえば、深刻な限界を含まない研究、深刻な限界を含む研究、非常に深刻な限界を含む研究が存在する場合に、深刻な限界の平均的評価に基づき、エビデンスの質を自動的に1段階グレードダウンしてはならない)。むしろ、質の高い研究に注目することを基本方針とし(後述のとおり)、各研究がエビデンスの質にどう影響するのかを慎重に検討するものとする。

第二に、慎重な検討を行うにあたっては、推定される効果の大きさに対し、それぞれの試験がどの程度寄与するのかについて、評価する必要がある。通常、各試験の寄与度は、研究のサンプルサイズとアウトカムイベントの数によって決まる。つまり、イベント数の多い大規模な試験は寄与度が大きく、これよりもさらにイベントの数の多いより大規模な試験では、寄与度がさらに大きくなる。

第三に、グレードダウンの判定は慎重に行うべきである。つまり、risk of bias(について評価を下げる際は、入手可能な一連のエビデンスの大部分において相当なrisk of biasが存在することに確信を持つべきである。

第四に、risk of biasは他の限界と照らし合わせて勘案すべきである。たとえばレビューアが、質に関わる2つの問題(risk of biasと精確さなど)について難しい判断を迫られる場合には、少なくともそのいずれか1つの問題について、グレードダウンを行うとよい。

第五に、前述の4つの指針にかかわらず、レビューアは難しい判断を迫られる状況に直面することになるだろう。この場合、レビューアはそのような状況にあることを認識するとともに、なぜそのような状況にあると考えるのかについて具体的に説明し、最終的な判断の根拠を明確に示すべきである。

■ Item #-14: メタ解析をする(メタ解析が不可能な場合もある)

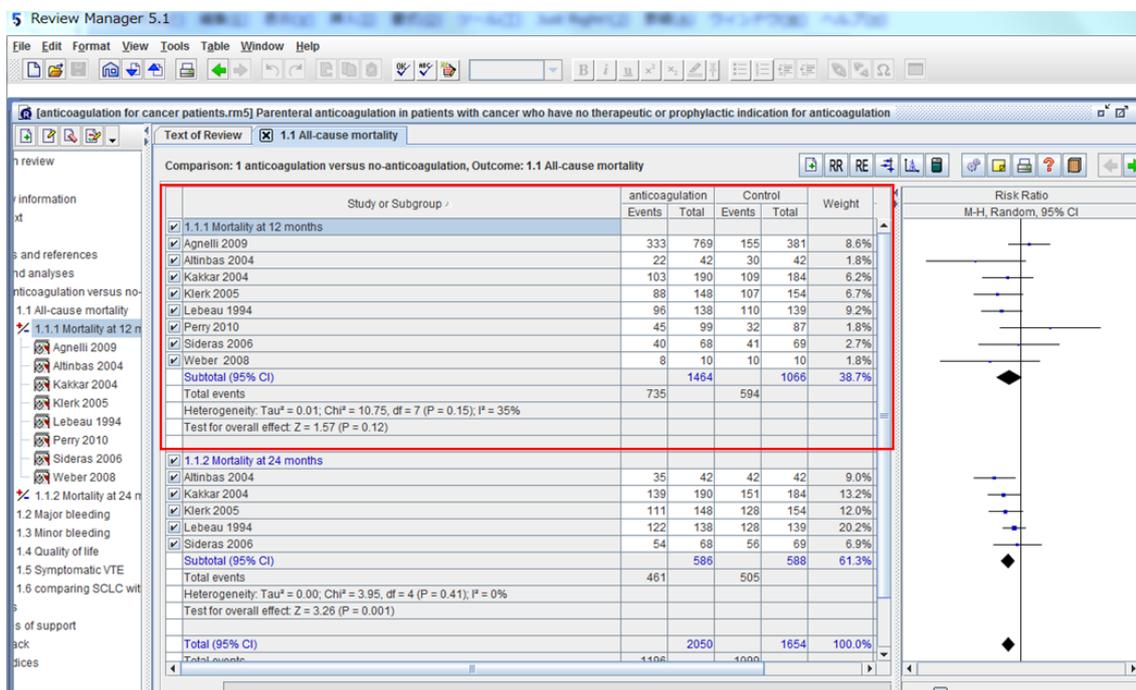
Blog: 2012/07/24

メタアナリシス(とは、過去に行われた複数の研究結果を定量的に統合し分析する統計学的な解析の方法の1つであるが、複数の臨床試験データを単純に平均したものではなく、データのばらつき具合で重み付けしてから統合したものである。統合の統計学的手法は、データの種類(2値、連続)、統合する統計値(オッズ比、相対リスク、リスク比、相関係数、加重平均差など)、統合する統計手法(固定効果モデル、ランダム効果モデル)などによって異なる。

1つの研究ではなく複数の研究を統合したものを総体エビデンス(body of evidence)と呼び、GRADE システムにおいては総体エビデンスの質を評価するのが基本である。しかし、利用できる最良のエビデンスが単一研究の場合もありえる。

コクラン共同計画では、選択論文からのデータは、RevManという統計解析ツールを用いてメタアナリシスを行っている。RevManでは、種々の異なった解析方法による結果の表示や、出版バイアスのチェックとしてのファンネルプロットも表示できる。さらに、RevManではGRADEプロットのデータ交換も可能である。

RevMan使用：アウトカム：全死亡率(12ヶ月)



参考

1. システムティック・レビューとメタアナリシス： GRADE 教科書 p37～39
2. メタアナリシス： [http://homepage3.nifty.com/aihara/meta\\_analysis3.html](http://homepage3.nifty.com/aihara/meta_analysis3.html)

附：メタアナリシスのためのソフトウェア

ソフトウェア	URL	入手
Review Manager (RevMan)	<a href="#">Cochrane Collaboration</a>	無料
Stata	<a href="#">STATA</a>	有料
SAS	<a href="#">SAS</a>	有料
Comprehensive Meta-analysis	<a href="#">Comprehensive Meta-analysis</a>	有料
MetaWin	<a href="#">MetaWin</a>	有料
Meta-DiSc	<a href="#">MetaDisc</a>	無料
MetaAnalyst	<a href="#">MetaAnalyst (β)</a>	無料
DistilerSR	<a href="#">DistilerSR</a>	有料
EPPI-Reviewer	<a href="#">EPPI centre</a>	有料
JBI SUMARI	<a href="#">Evidence-Based Practice Network</a>	有料
MIX	<a href="#">MIX 2.0</a>	有料
WINBUGS	<a href="#">BUGS project</a>	有料
メタアナリシス計算表	<a href="#">Microsoft Excel(森實敏夫)</a>	無料
R (S-PLUS)	<a href="#">メタ・アナリシス入門(丹後敏郎)</a>	有料
Meta-analysis	<a href="#">Microsoft Excel (BMC Research Notes)</a>	無料

## ■ Item #-15: エビデンスの質: 研究デザイン

Blog: 2012/07/25

GRADEにおいては、研究ごとにエビデンスを評価するのではなく、複数の研究のシステマティック・レビューにより、アウトカムを主体として横断的に収集した総体エビデンス(body of evidence)の質を、アウトカムごとに評価する。

### エビデンスの質とは:

システマティック・レビューにおけるエビデンスの質とは、「ある効果推定値(estimate of effect)が正しいという確信(confidence)が、どの程度か」と定義される。システマティック・レビューでは推奨を作成しないため、各アウトカムについてエビデンスの質のグレーディングを行う。

([GRADEシステムによるエビデンスの質評価と定義](#)、参照)

一方、診療ガイドラインにおけるエビデンスの質は、「その効果推定値に対する確信が、ある特定の推奨を支持する上でどの程度十分か」を示す。GRADE では、“患者にとって重大な各アウトカム”に関するエビデンスの質を個別に評価し、それらを全体として見渡して、アウトカム全般に関する全体的なエビデンスの質(overall quality of evidence)を1つだけ決定する。

### エビデンスの質と研究デザイン

エビデンスの質の評価では、最初に研究デザインの評価をする。ランダム化比較試験(RCT)のエビデンスの質は「高」、RCT以外の観察研究のエビデンスの質は「低」とする(表1.4-1)。

シナリオの場合、コクランレビューに組み込まれた9件の試験の研究デザインは、ランダム化比較試験であり、エビデンスの質は「高」である。

(スライド #33)

注: (GRADE教科書 p42)

1. 観察研究の質を評価するとき、前向き研究と後ろ向き研究の区別をすることはあまり意味がない。
2. 専門家の意見はGRADEではエビデンスの分類には含まない。

附：GRADEシステムによるエビデンスの質評価と定義

表 1.4-1 GRADE エビデンスの質評価と定義

エビデンスの質 (Quality of evidence)	研究デザイン (Study design)	グレードを下げる (*) (Lower if)	グレードを上げる (*) (Higher if)
高 high	RCT	<ul style="list-style-type: none"> <li>研究の限界 (limitations)                             <ul style="list-style-type: none"> <li>-1 深刻 (serious)</li> <li>-2 非常に深刻 (very serious)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>効果の程度が大きい (large magnitude of effect)                             <ul style="list-style-type: none"> <li>+1 大きな効果 RR &gt; 2 あるいは &lt; 0.5</li> </ul> </li> </ul>
中 moderate			
低 low	観察研究	<ul style="list-style-type: none"> <li>結果の非一貫性 (inconsistency)                             <ul style="list-style-type: none"> <li>-1 深刻</li> <li>-2 非常に深刻</li> </ul> </li> <li>エビデンスの非直接性 (indirectness)                             <ul style="list-style-type: none"> <li>-1 深刻</li> <li>-2 非常に深刻</li> </ul> </li> <li>データの不精確さ (imprecision)                             <ul style="list-style-type: none"> <li>-1 深刻</li> <li>-2 非常に深刻</li> </ul> </li> <li>出版バイアス (publication bias)                             <ul style="list-style-type: none"> <li>-1 ありそう (likely)</li> <li>-2 非常にありそう (very likely)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>+2 極めて大きい効果 RR &gt; 5 あるいは &lt; 0.2</li> <li>用量-反応勾配 (dose-dependent gradient)                             <ul style="list-style-type: none"> <li>+1 あり</li> </ul> </li> <li>すべての交絡因子 (plausible confounder)                             <ul style="list-style-type: none"> <li>+1 提示された効果を減弱させている</li> </ul> </li> </ul>
非常に低 very low			
* 1 = 1 段階グレードを下げる, あるいは上げる (たとえば, 「高」 から 「中」 へ) 2 = 2 段階グレードを下げる, あるいは上げる (たとえば, 「高」 から 「低」 へ)			
アウトカムについての研究全般に関するエビデンスの質の定義 (Definition: Quality of evidence across studies for the outcome)			
<ul style="list-style-type: none"> <li>「高」 = 今後の研究によって効果推定値に対する確信性が変わる可能性は低い (very unlikely)。</li> <li>「中」 = 今後の研究によって効果推定値に対する確信性に重要な影響がおよぶ可能性が高く (likely), 推定値が変わる可能性がある (may)。</li> <li>「低」 = 今後の研究によって効果推定値に対する確信性に重要な影響がおよぶ可能性が非常に高く (very likely), 推定値が変わる可能性が高い (likely)。</li> <li>「非常に低」 = あらゆる効果推定値が不確実である (very uncertain)。</li> </ul>			

5つのグレードダウン、3つのグレードアップ要因があるが、各評価を定量的に行ってはいけない。-1と-1が2つ存在したから、必ず2段階下げるといってはいけない。

■ Item #-16: エビデンスの質：(グレードダウンの5要因) limitations (risk of bias)

Blog: 2012/07/26

総体エビデンス(body of evidence)についてグレードダウンの5要因の1つであるrisk of biasを評価する際は、”コクランrisk of bias評価をGRADE の「研究の限界(limitations)」評価へ直接的に反映させる”。

両者の対応は、「low risk of bias」は「no limitations (限界なし)」、「unclear risk of bias (不明)」は「no serious (限界なし)」または「serious (深刻)な限界あり」、「high risk of bias」は「serious (深刻)な限界」または「very serious (非常に深刻)な限界」に相当する。

GRADE 教科書 p51

表 1.5.1-6 コクラン risk of bias 分類と GRADE の「研究の限界」 分類の対応

コクラン Risk of bias	GRADE の「研究の限界 (Limitations)」
Low risk of bias	No serious limitations
Unclear risk of bias	No serious limitations or serious limitations
High risk of bias	Serious or very serious limitations

表 1.5.1-7 RCT における「研究の限界」 (risk of bias) の評価 <sup>48,83)</sup>

Risk of bias / 限界 (limitations)	個別研究 (Within a study)	複数の研究全般 (Across studies)	複数の研究全般に関する解釈* (Interpretation across studies)
No serious limitations 深刻な限界はなく、グレードは下げない	すべての主要基準 (表 1.5.1-2, 47 頁) に関して, Low risk of bias である。	ほとんどの情報が「Low risk of bias」研究から得られている。	エビデンスの質「高」: 真の効果が効果推定値の近くに存在する。
Serious limitations 深刻な限界があり 1 段階下げる (「高」から「中」)	1 つの基準項目についての極めて重要な限界 (Crucial limitations), あるいは、いくつかの基準での限界があり、そのために効果推定値の確信を低下させるのに十分 (Sufficient) である。	ほとんどの情報が「Moderate risk of bias」 <sup>5</sup> 研究から得られている。	エビデンスの質「中」: 真の効果が効果推定値に近いと考えられるが、大幅に異なる可能性もある。
Very serious limitations 非常に深刻な限界があり、 2 段階下げる (「高」から「低」)	1 つあるいはそれ以上の基準に極めて重要な限界があり、そのために明らかに効果推定値の確信を低下させる。	ほとんどの情報が「High risk of bias」研究から得られている。	エビデンスの質「低」: 真の効果が効果推定値とは大きく異なるかもしれない。

注：個々の研究のコクラン risk of bias テーブルに基づき、複数の研究にわたる全体的なエビデンスの質(overall quality of evidence across studies) を評価するための指針 (GRADE working group: JCE) → スライド#42

シナリオにおける、2つのアウトカムに関する評価:

- 全死亡率: no serious risk of bias (スライド #44)
- 大出血: no serious risk of bias

■ Item #-17: エビデンスの質:(グレードダウンの5要因) inconsistency

Blog: 2012/07/27

非一貫性あるいは不一致性 (inconsistency) とは、研究によって治療効果の推定値が大きく異なる(すなわち、結果に異質性(heterogeneity)または、ばらつきが存在する)ことを指し、根本的な治療効果に真の差異が存在することを意味する。異質性があるにもかかわらず研究者が妥当な説明ができない場合、その程度に応じてエビデンスの質を1～2段階下げる。

表 1.5.2-2 非一貫性の4つの基準

- |  |
|--|
| <ol style="list-style-type: none"><li>1. 点推定値が研究間で異なり、その相違がかなり大きい。</li><li>2. 各信頼区間の重なりが、ほとんどまたは全くない。</li><li>3. 異質性の検定 (二択の帰無仮説：全研究で差がない) で、<math>p</math> 値が小さい。</li><li>4. <math>I^2</math> 値 (研究間の異質性を示す) が高い。</li></ol> |
|--|

(GRADE 教科書 p52)

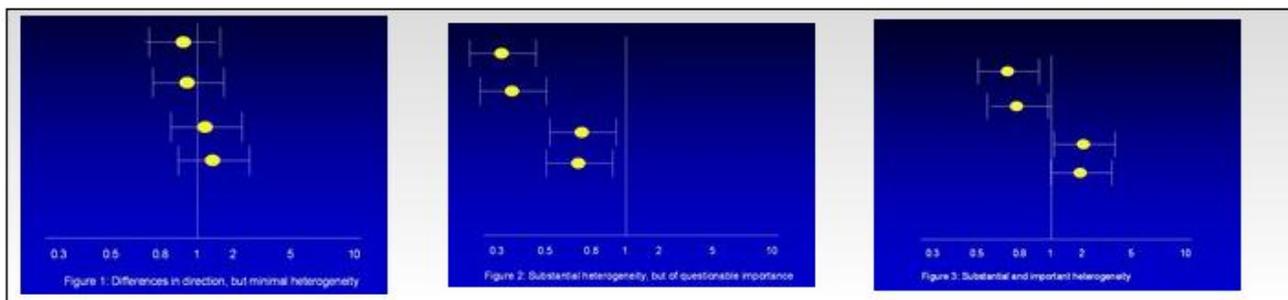
効果の方向は非一貫性の基準ではない

GRADE handbook: 「説明のつかない結果の異質性または非一貫性について」を参照

<http://homepage3.nifty.com/aihara/GRADEproHelp/Inconsistent%20results.htm>

<http://www.grade-jpn.com/grade/guideline/inconsistency.html>

inconsistency と重要性とは別問題ということに注意すべきである。ガイドラインパネルは、inconsistency が大きいとしても、それを重要なものではないと判定する場合もある。(参照一下図2)

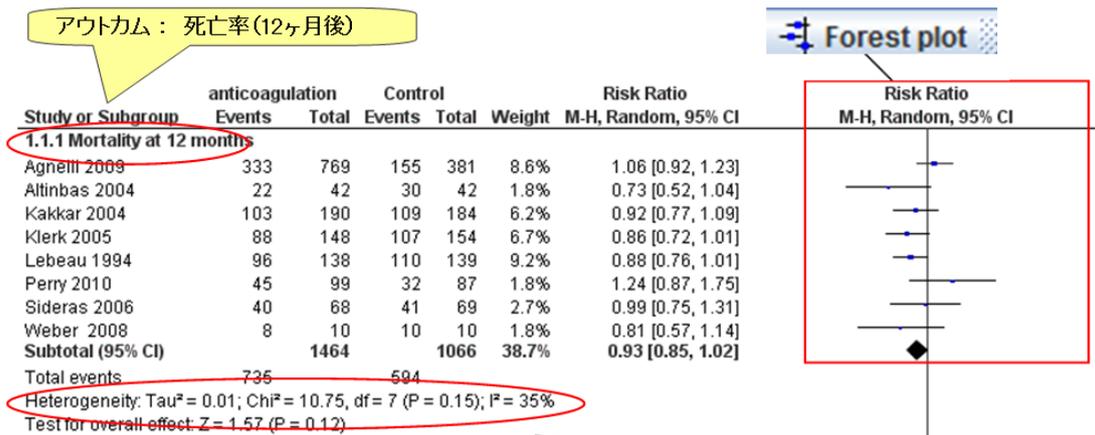


(本図は、JCE 2012 論文にて発表されているものと同じである)

シナリオにおける、2つのアウトカムに関する評価：

- 全死亡率： **serious inconsistency (-1)** (スライド #49)
- 大出血： **no serious inconsistency**

RevManの forest plotをみる



点推定値のばらつきがややあるが、信頼区間はほとんどの研究で重なっている。異質性検定は  $p > 0.05$ 、さらに研究間の異質性を示す  $I^2$  値は35% (中等度) である。  
 本アウトカムに関する結果の異質性は、(imprecisionと合わせて)、エビデンスの質を1段階下げる。

■ Item #-18: エビデンスの質：(グレードダウンの5要因) indirectness

Blog: 2012/07/30

エビデンスの非直接性(indirectness)は、従来の「外的妥当性」(external validity)(一般化可能性generalizability と同義)と同様のものである。Indirectnessには2つのタイプがある。

非直接性には2つのタイプがある

1. 非直接的な比較(\*) - 介入 A と B の比較ができないために、A を C と比較し、B を C と比較したときがこれに当てはまる。このような研究では、A 対 B の効果の大きさを非直接的に比較できる。しかし、このようなエビデンスは、A と B の直接的比較が提供する質よりも低いものである。
2. 非直接的な集団、介入、比較、またはアウトカム - ガイドラインパネルまたはシステムティック・レビューの著者が取り組む疑問が、集団、介入、比較、またはアウトカムに関する入手可能なエビデンスと違っている。

(\*) 非直接的な比較 indirect comparison

例； ネットワークメタアナリシス(NMA)：

個々の試験に発表されている効果推定値とその信頼区間を利用して、論文の中で直接比較していない治療群間の差を推定する方法(mixed treatment comparison, multiple treatments meta-analysis ともいう)。通常は、NMA のエビデンスの質は indirectness のために、1～2段階グレードダウンになる。

注意：

Surrogate outcomes の全てがエビデンスの質低下ではない<sup>2</sup>

Rarely, surrogates are sufficiently well established that review authors or guideline panelists should choose not to rate down quality of evidence for indirectness. In our view, this should be restricted to situations in which, within the same class of drug (e.g., beta-blockers, calcium antagonists, diuretics, bisphosphonates), changes in the surrogate have repeatedly proved closely related to changes in the patient-important outcome in the context of RCTs.

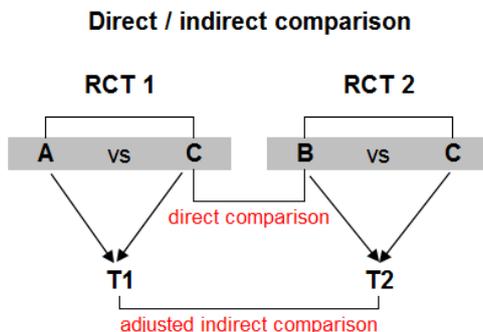
シナリオにおける、2つのアウトカムに関する評価：

- ・全死亡率： no serious indirectness
- ・大出血： no serious indirectness

<sup>2</sup> GRADE guidelines: 8. Rating the quality of evidence—indirectness  
Journal of Clinical Epidemiology Volume 64, Issue 12, Pages 1303-1310, December 2011

附：indirect comparison

介入 A と介入 B の効果を比較したいが、A と B を直接比較した RCT がない場合、A 対 C 研究と B 対 C 研究を比較解析するものが、indirect comparison である。



この場合、A と B を直接比較するのは (naive method) は統計的に問題があり、adjusted indirect comparison を使わないといけない。adjusted indirect comparison の結果は、head to head RCT とよく一致するという報告もある。adjusted indirect comparison は、メタアナリシスのサブグループ解析にも適用すべきである。

Adjusted indirect method: 文献において使われる他の用語
<ul style="list-style-type: none"> <li>・Cross-trial comparison</li> <li>・Connected comparative experiment</li> <li>・Network meta-analysis</li> <li>・Mixed comparison</li> <li>・Virtual comparison</li> <li>・Multiple treatments meta-analysis</li> </ul>
注意
<ul style="list-style-type: none"> <li>・When available, results from direct and indirect comparisons usually agree but they may not</li> <li>・Essential to consider how similar (sets of ) trials are: study and patient characteristics</li> <li>・Indirect and direct data should always be considered separately and not automatically pooled</li> <li>・Interpretation should always be cautious in view of the observational nature of the data</li> </ul>

Altman DG: The issue of indirect comparisons in meta-analysis, 京都シンポジウム(2005)を改変

■ **indirect comparison** についての GRADE アプローチの基本は以下である。

1. 直接エビデンスをグレードする。
2. 各 loop の間接エビデンスをグレードする。
3. 最も質の高いエビデンス(直接 and/or 間接・loop)を使って点推定値を計算する。
4. 推定値をグレードする。

■ Item #-19: エビデンスの質：(グレードダウンの5要因) imprecision

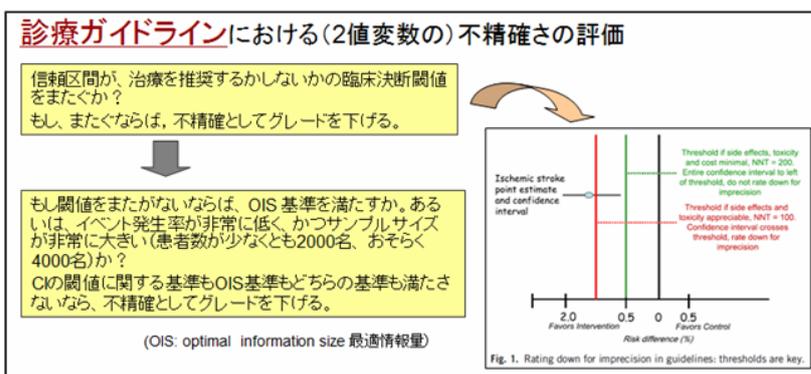
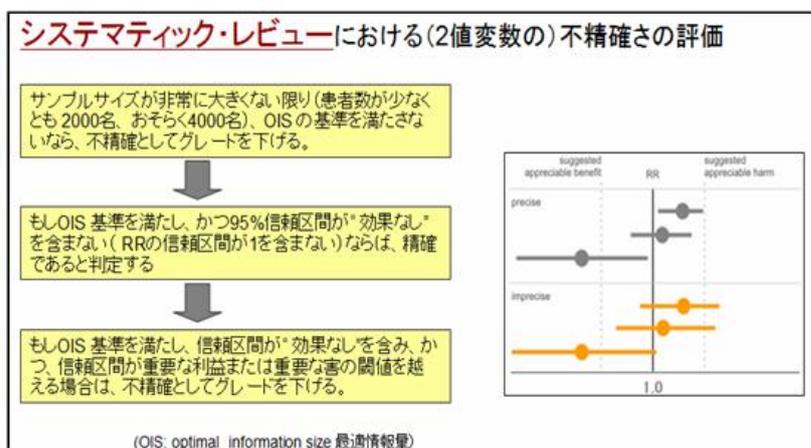
Blog: 2012/08/02

データの不精確さ(imprecision)とは、データのばらつきが大きい場合や対象患者数(サンプルサイズ)やイベント数が少ない場合に信頼区間の幅が広いことを指す。

GRADE では、システマティック・レビューのエビデンスの質の定義と診療ガイドラインのエビデンスの質の定義を区別しているため、不精確さを理由にエビデンスの質評価のグレードを下げる基準もシステマティック・レビューと診療ガイドラインでは異なる。

参考

1. データの不精確さ (imprecision): GRADE 教科書 p56-59
2. スライド #57、#60



シナリオにおける、2つのアウトカムに関する評価：

- ・全死亡率： no serious imprecision(\*)
- ・大出血： no serious imprecision

(\*)実際のコクランレビューにおいては、Inconsistency と imprecision の判定で、(2段階ではなく)1段階のグレードダウンとしている。

■ Item #-20: エビデンスの質：(グレードダウンの5要因) publication bias

Blog: 2012/08/03

出版バイアスは、研究が選択により偏って出版されることが原因で本来のプラス効果またはマイナス効果が系統的に過小または過大に評価されることをいう。

出版バイアスに対処するための一般的な4つの方法が報告されている。

(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)

表 20.1-3

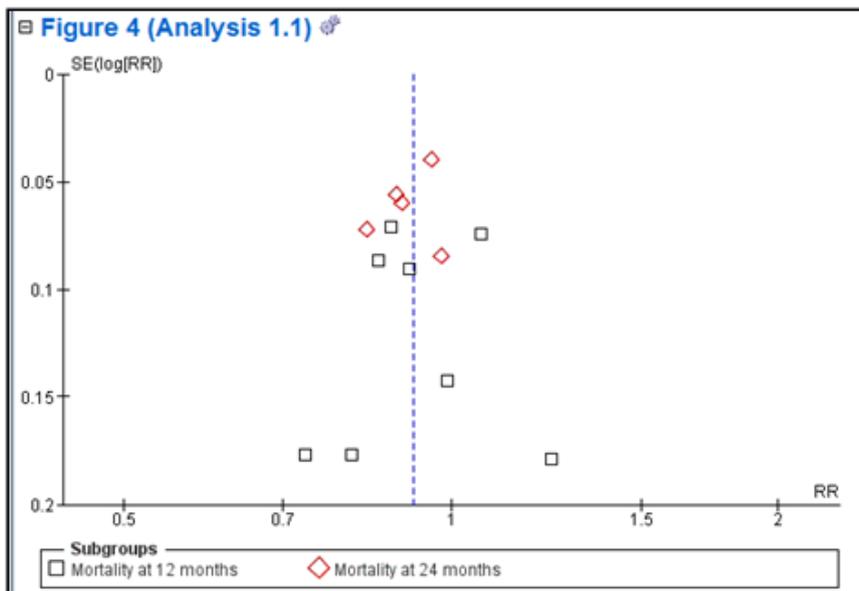
報告バイアスに対処するための4つの手法

1. 小規模研究によってより大きな効果が示されていないか調べる
  - a. ファンネルプロットの目視評価
  - b. ファンネルプロットの統計的解析
2. 想定される出版バイアスを把握した後、関式を復元することによって、エビデンスを再構築する
  - a. trim and fill
  - b. フェイルセーフ数 Fail-safe N
3. 統計的有意水準に応じた、出版の確率を推定する
4. データの出現に応じて経時的に効果サイズの変化を調べる

注：厳密な意味では、“報告バイアス(reporting bias)”、特に選択的アウトカム報告バイアスは、GRADEの limitations(risk of bias)に含まれる。

シナリオにおける、2つのアウトカムに関する評価：

- ・全死亡率： **undetected**
- ・大出血： **undetected**



(ファンネルプロット:RevMan を使って作成、相原)

■ Item #-21: エビデンスの質:(グレードアップの3要因) 効果の程度が大きい

Blog: 2012/08/06

関連性(association)を示す効果の大きさの指標として、GRADEでは原則として相対リスク(relative risk: RR)を用いる。

表 1.6.1-1 効果の大きさとエビデンスの質を上げる程度

効果の大きさ (Magnitude of effect)	効果の指標 (Effect measure)	エビデンスの質 (Quality of evidence)
大きい (Large)	RR > 2 または < 0.5	1 段階評価を上げる
非常に大きい (Very large)	RR > 5 または < 0.2	2 段階評価を上げる

(RR: 相対リスク)

(GRADE 教科書 p63)

参考: 大きい効果の程度について(GRADE ハンドブック)

<http://homepage3.nifty.com/aihara/GRADEproHelp/Large%20or%20very%20large%20effect.htm>

評価例: シナリオにおける、2つのアウトカムに関する評価:

- ・全死亡率: **大きな効果はない** (RR 0.93)
- ・大出血: **大きな効果はない** (RR 1.3)

注:

1. グレードアップ要因は、グレードダウンの要因と切り離して評価するものではない。すなわち、limitations(RoB)が深刻なもの、あるいは imprecision が非常に深刻なもので、エビデンスのレベルが下がったものを、“効果の程度が大きい”としてグレードを上げるようなことはしない。

2. 「関連性(association)がある」と「因果関係(causality または causation)がある」ことは異なることに注意する必要がある。

関連性の評価において従来利用されてきた **Bradford Hill's criteria** は、GRADEシステムによって説明可能であると発表されている。(スライド # 67)

[The GRADE approach and Bradford Hill's criteria for causation  
JECH 2011;65:392-395 doi:10.1136/jech.2010.119933](#)

■ Item #-22: エビデンスの質:(グレードアップ3要因) 交絡因子

Blog: 2012/08/07

なんらかの理由で評価を下げられていない観察研究において交絡因子(confounding factor または confounder)があると考えられる場合に、「真の効果はあるのに、すべての交絡因子が報告された効果を減少させる方向に働く」、あるいは「真の効果はないのに、すべての交絡因子が効果を増大させる方向に働く」ことがある。

真の効果が存在するのに、考えられうるすべての交絡因子が同じ方向に働くことで見かけ上の介入効果が真の効果よりも弱められていると考えられる場合、エビデンスの質を上げる。

たとえば、もし、対照群よりも治療群により重病の患者が多いにもかかわらず、治療群でより死亡率が減少したということであれば、それは実際の新規治療の効果が得られた結果以上であることを示唆している。このため、他に妥当性についての問題がない場合は、エビデンスの質を1段階上げる。

参考: 全てのありそうな交絡の影響について(GRADE ハンドブック)

<http://homepage3.nifty.com/aihara/GRADEproHelp/Plausible%20biases.htm>

評価例: シナリオにおける、2つのアウトカムに関する評価:

- 全死亡率: 評価なし
- 大出血: 評価なし

■ Item #-23: エビデンスの質：(グレードアップ3要因) 用量反応勾配

Blog: 2012/08/08

複数の観察研究から得られた結果において、用量反応勾配(dose-response gradient)がある場合は、エビデンスの質を1段階上げる。ただし、評価を上げることができるのは、研究の妥当性に問題がなく、評価が下げられていない観察研究に限られる。

たとえば、1日の喫煙量の増加と肺癌による死亡リスクの増加という関連性は、喫煙が肺癌を引き起こすというエビデンスの質を上げる理由となる。

また、ワルファリンによる抗凝固療法を受けている患者において、抗凝固の程度の指標であるINR(国際標準比)の高値と出血リスク増加との間に用量反応勾配があるという観察は、治療域を超えた抗凝固レベルが出血リスクを増加させることの確信を高めるものである。

参考：用量反応勾配について(GRADE ハンドブック)

<http://homepage3.nifty.com/aihara/GRADEproHelp/Dose-response%20gradient.htm>

評価例：シナリオにおける、2つのアウトカムに関する評価：

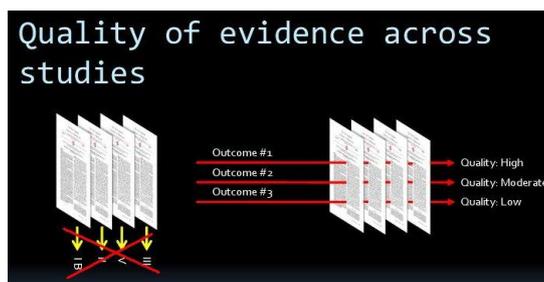
- ・全死亡率： 評価なし
- ・大出血： 評価なし

理論的にはランダム化比較試験の結果についてグレードアップすることは可能であるが説得力のある適切な報告例はなく、通常はグレードアップの3要因は観察研究を対象としたものである。(JCE #9: Rating up the quality of evidence)

■ Item #-24: 各アウトカムについて、総体エビデンスの質(body of evidence)を判定する

Blog: 2012/08/09

GRADEにおいては、個々の研究のエビデンスを評価するのではなく、複数の研究のシステマティック・レビューを使う。すなわち、アウトカムを主体として研究横断的に収集した総体エビデンス(body of evidence)の質を、アウトカムごとに評価する。



グレードダウン要因(例、imprecision と risk of bias)について、-1と-1が2つ存在したから、必ず2段階下げるということでなく、検討の結果1段階のみ下げるとすることもある(GRADEproでは、エビデンステーブルの作成があるので、より深刻な方のみを-1として調整する)。これらの判断のポイントを脚注(\*)に記述することが重要である。

(\*) [http://www.grade-jpn.com/anticoag\\_cancer-EP.pdf](http://www.grade-jpn.com/anticoag_cancer-EP.pdf)

(本エビデンスプロファイルは、2010年3月に作成したものである。CQは本シナリオのものと同じであるが、利用しているレビューは改訂前の2008年発表のCDSRである。その後のRCT研究の追加により改訂CDSRにおいてはエビデンスの質が低下している。)

シナリオにおける、5つのアウトカムに関する QoE 評価:

(スライド #78)

アウトカム	エビデンスの質
全死亡率	Moderate (B) ⊕⊕⊕○
症候性血栓塞栓症	High (A) ⊕⊕⊕⊕
大出血	Moderate (B) ⊕⊕⊕○
小出血	Moderate (B) ⊕⊕⊕○
HQOL	Low (C) ⊕⊕○○

■ Item #-25: エビデンスプロファイル作成、SoF 表作成

Blog: 2012/08/10

GRADEエビデンステーブルには、エビデンスプロファイル、SoF (Summary of findings)テーブル、レビュー概要テーブルがあり、そのうち、前2者が代表的である。エビデンステーブルには、最も質の高いエビデンスを提示すべきである。

[1] GRADE エビデンスプロファイルは、各アウトカムに関するエビデンスの質評価の詳細や研究の結果要約とともに、その判断理由などを脚注に記載した表である。

シナリオのエビデンスプロファイル (スライド #71 参照)

[http://homepage3.nifty.com/aihara/EP\\_heparin\\_accp\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/EP_heparin_accp_aihara_20120426.html)

## GRADE evidence profile

[http://homepage3.nifty.com/aihara/EP\\_heparin\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/EP_heparin_aihara_20120426.html)

Author(s): aihara E, Aki A, Sameer K, Gunukula, Frederiek F van Doormaal, Maddalena Barba, Saskia Kuipers, Saskia Middeldorp, Victor E O Yosuico, Heather O Dickinson, Holger Schunemann  
 Date: 2012-04-17  
 Question: Should heparin vs no heparin be used for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation?  
 Settings: outpatient  
 Bibliography: Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation Aki EA, Gunukula M, Barba VED, Yosuico FF, van Doormaal S, Kuipers S, Middeldorp HO, Dickinson A, Bryant H, Schunemann. Cochrane Database Syst Rev. 2011 Apr 13;(4):CD006652.

No of studies	Design	Risk of bias	Quality assessment				No of patients	Effect	Quality	Importance
			Inconsistency	Indirectness	Imprecision	Other considerations				
8	randomised trials	no serious risk of bias <sup>1</sup>	serious <sup>2</sup>	no serious indirectness	no serious imprecision <sup>3</sup>	none	735/1464 (50.2%) 594/1066 (55.7%) 64.5% 79.1%	RR 0.93 (0.85 to 1.02) 39 fewer per 1000 (from 84 fewer to 11 more) 45 fewer per 1000 (from 97 fewer to 13 more) 55 fewer per 1000 (from 119 fewer to 16 more)	⊕⊕⊕ MODERATE	CRITICAL
All-cause mortality (follow-up 12 months)										
7	randomised trials	no serious risk of bias <sup>1</sup>	no serious inconsistency	no serious indirectness	no serious imprecision	none	38/1338 (2.8%) 57/926 (6.2%) 2.9%	RR 0.55 (0.37 to 0.82) 28 fewer per 1000 (from 11 fewer to 39 fewer) 13 fewer per 1000 (from 5 fewer to 18 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Symptomatic VTE (follow-up 12 months)										
9	randomised trials	no serious risk of bias <sup>1</sup>	no serious inconsistency	no serious indirectness	serious <sup>4</sup>	none	30/1624 (1.8%) 23/1219 (1.9%) 0.7%	RR 1.3 (0.59 to 2.88) 6 more per 1000 (from 8 fewer to 35 more) 2 more per 1000 (from 3 fewer to 13 more)	⊕⊕⊕ MODERATE	CRITICAL
Major bleeding (follow-up 12 months)										

アウトカム

質評価

(Continued)

[2] SoF テーブルは、完全なエビデンスプロファイルと同じ情報からエビデンスの質評価の詳細を省略し、コメント欄を追加したものである。標準的なSoFテーブルには、7項目 (アウトカム、対照群リスクあるいは想定リスク、介入群リスクあるいは対応リスク、効果の相対的な大きさ、参加者数と研究件数、アウトカムごとの全体的なエビデンスの質、コメント：列順は左から重要度を反映し、重要な列から表示) が含まれるよう推奨されているが、エビデンスプロファイルと同様に表示様式については画一的なものはない。

(GRADE教科書 第2章: GRADEシステムのエビデンステーブルを参照)

シナリオの SoF テーブル (スライド #74 参照)

[http://homepage3.nifty.com/aihara/sof\\_heparin\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/sof_heparin_aihara_20120426.html)

## Summary of finding (SoF)テーブル

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	no heparin	heparin				
<b>Mortality</b> Follow-up: 12 months	Medium risk population 649 per 1000	664 per 1000 (552 to 662)	RR 0.93 (0.85 to 1.02)	2531 (8 studies)	⊕⊕⊕○ moderate <sup>1,3</sup>	
<b>Symptomatic VTE</b> Follow-up: 12 months	Medium risk population 29 per 1000	16 per 1000 (11 to 24)	RR 0.55 (0.37 to 0.82)	2264 (7 studies)	⊕⊕⊕⊕ high <sup>1</sup>	
<b>Major bleeding</b> Follow-up: 12 months	Medium risk population 7 per 1000	9 per 1000 (4 to 20)	RR 1.3 (0.59 to 2.88)	2843 (9 studies)	⊕⊕⊕○ moderate <sup>1,4</sup>	
<b>Minor bleeding</b> Follow-up: 12 weeks	Medium risk population 27 per 1000	28 per 1000	RR 1.05 (0.75 to 1.46)	2345 (7 studies)	⊕⊕⊕○ moderate <sup>1,4</sup>	

SoFiに関するFAQの翻訳 (\*)

[http://www.grade-jpn.com/grade\\_pro-faq.html](http://www.grade-jpn.com/grade_pro-faq.html)

[http://homepage3.nifty.com/aihara/sof\\_heparin\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/sof_heparin_aihara_20120426.html)

(Continued)

注:

エビデンスプロファイルや SoF テーブル内のデータ解釈については、GRADE 教科書の第 2 章、あるいは「GRADE システムの使い方」(ppt)のスライド # 76、77などを参照のこと。

### SoF に関する FAQ

- Q-1: SoF テーブル(SoF table)を理解する必要があるのは誰か?
- Q-2: レビューの SoF テーブルの作成責任は誰にあるのか?
- Q-3: SoF テーブルの作成後は誰にチェックをしてもらえるのか?
- Q-4: レビューの SoF テーブルの作成に向けた構想はどのタイミングで考え始めるのか?
- Q-5: SoF テーブルに 7 件以上のアウトカムを含めたい場合はどうしたらいいのか?
- Q-6: SoF テーブルはどのようにして作成するのか?
- Q-7: GRADEpro を使って SoF テーブルを作成する場合は何をしなければならないのか?
- Q-8: Risk of Bias テーブルと SoF テーブルの間にはなんらかの関連性があるか?
- Q-9: エビデンスの評価システムが多数存在する中で、なぜ GRADE を使用するべきなのか?
- Q-10: システマティック・レビューにおいて GRADE アプローチを使ってエビデンスの質を評価する際に特別なスキルが必要か?

■スライド 76

# 対照群リスクからの対応リスクの計算

Study or Subgroup	anticoagulation		Control		Weight	Risk Ratio		Risk Ratio M-H, Random, 95% CI
	Events	Total	Events	Total		M-H, Random, 95% CI	M-H, Random, 95% CI	
1.1.1 Mortality at 12 months								
Agnelli 2009	333	769	155	381	8.6%	1.06 [0.92, 1.23]		
Altinbas 2004	22	42	30	42	1.8%	0.73 [0.52, 1.04]		
Kakkar 2004	103	190	109	184	6.2%	0.92 [0.77, 1.09]		
Klerk 2005	88	148	107	154	6.7%	0.86 [0.72, 1.01]		
Lebeau 1994	96	138	110	139	9.2%	0.88 [0.76, 1.01]		
Perry 2010	45	99	32	87	1.8%	1.24 [0.87, 1.75]		
Sideras 2006	40	68	41	69	2.7%	0.99 [0.75, 1.31]		
Weber 2008	8	10	10	10	1.8%	0.81 [0.57, 1.14]		
Subtotal (95% CI)		1464		1066	38.7%	0.93 [0.85, 1.02]		
Total events	735		594					
Heterogeneity: Tau <sup>2</sup> = 0.01; Chi <sup>2</sup> = 10.75, df = 7 (P = 0.15); I <sup>2</sup> = 35%								
Test for overall effect: Z = 1.57 (P = 0.12)								

median risk (RevMan5を使ったメタアナリシスからGRADEproにより自動的に計算)

Outcome: All-cause mortality

Number of p with event: Intervention: 735 total: 1464 50.2 %

with event: Control: 594 total: 1066 55.7 %

Control risk: Low Moderate 64.5 % High 79.1 %

Estimate of the effect: Relative: RR 0.93 of 0.93 95% CI from 0.85 to 1.02

Absolute: 39 fewer per 1000 95% CI from 84 to 11

対照群リスク 55.7%に対応した絶対効果が1000人あたりの数として提示されている

■スライド 77

# 対照群リスクからの対応リスクの計算

SoF: [http://homepage3.nifty.com/aihara/sof\\_heparin\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/sof_heparin_aihara_20120426.html)

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)
	Assumed risk	Corresponding risk			
All-cause mortality Follow-up: 12 months	No heparin	Heparin			
	Study population		RR 0.93	2531	⊕⊕⊕⊕
	557 per 1000	518 per 1000 (473 to 568)	(0.85 to 1.02)	(8 studies)	moderate <sup>1,2,3</sup>
	Moderate	645 per 1000 (548 to 658)			
	High	791 per 1000 (672 to 807)			

計算例:  
 $RR (0.93) \times 645 = 548$   
 $Lower\ CI (0.85) \times 645 = 548$   
 $Upper\ CI (1.02) \times 649 = 658$

No of patients	No heparin	Relative (95% CI)	Effect		Quality	Importance
			Absolute			
735/1464 (50.2%)	594/1067 (55.7%)	RR 0.93 (0.85 to 1.02)	39 fewer per 1000 (from 84 fewer to 11 more)	⊕⊕⊕⊕ MODERATE	CRITICAL	
	64.5%	45 fewer per 1000 (from 97 fewer to 13 more)				
	79.1%	55 fewer per 1000 (from 119 fewer to 16 more)				

■ Item #-26: 医療資源に関する資料作成

Blog: 2012/08/16

GRADE ハンドブック (<http://homepage3.nifty.com/aihara/GRADEproHelp.html>) を参照。  
資源の利用(コスト)を推奨に盛り込むときは、特に次の点を検討する。

- [コストと他のアウトカムの違い](#)は何か。
- どの[立場](#)を取るのか。
- [資源の影響](#)としてどの影響を盛り込むのか。
- [エビデンスの質](#)をどう判断するのか。
- これらの影響をどのように[提示](#)するのか。
- [形式的経済学モデル](#)にはどのような潜在的有効性があるか。
- [推奨を作成する上で、資源の利用をどう検討するのか](#)。

ガイドラインパネルは、資源の利用を検討する場合、コストについて検討する前に、まず他のアウトカムに関するエビデンスの質を決定し、利益と不利益のバランスを検討する必要がある。資源の利用の重要度に関する意思決定は、この最初のステップを基に行うものである。たとえば、健康への正味の利益[net health benefits]に関するエビデンスが欠如している状況では、資源の影響について検討しても意味がない。また、介入の利益が不利益を大幅に上回る場合は、資源の利用を検討する意味は低くなる。通常、資源の利用が重要となるのは、利益と不利益が拮抗している場合である。

(\*) 資源利用(コスト)に関するエビデンスサマリー

バランスシート(エビデンスプロファイル)は、正味の利益が増分コストに見合うかどうか(net benefits are worth the incremental costs)の判断ができるものでなければならない。

2012年8月、JCEシリーズ [Grade guidelines: 10. Considering resource use and rating the quality of economic evidence](#) において、資源利用に関する GRADE 評価の考え方やバランスシートの長所短所、資源利用のエビデンスプロファイル、SoFなどが発表された。

■ GRADE を使って資源利用を検討する際のキーポイント

**Box 3 Key points in considering resource implications using the GRADE approach**

- Only important or critical resource use should be included in an evidence profile.
- Evidence must be found providing an estimate of the difference in resource use between the intervention and the comparison group.
- Resource use should be presented in natural units (e.g., days in hospital, minutes of clinician time).
- The quality of evidence should be appraised explicitly for each important or critical resource on-sequence using the same criteria as for health outcomes.

■ エビデンスプロファイル例

Table 1. Example of resource use evidence profile

Studies (follow-up)	Quality assessment						Summary of resources and costs			
	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other factors	No of patients	Resources costs per patient (1999 AU \$)		Overall quality
Drugs (6 mo) One study (Doran, 2003) <sup>a</sup>	RCT	No	No	Some uncertainty <sup>b</sup>	No	None	405	Resources (mean daily) 57 mg	11 mg	Moderate ⊕⊕⊕○
								Costs (6 mo) 37 (33 SD)	459 (461 SD)	
Other health care costs (6 mo) One study (Doran, 2003) <sup>c</sup>	RCT	No	No	Some uncertainty <sup>b</sup>	No	None	405	Resources NA	NA	Moderate ⊕⊕⊕○
								Costs (6 mo) 1,378 (NA)	1,270 (NA)	
Crime costs No information available <sup>d</sup>										

**Question:** Should buprenorphine maintenance flexible doses vs. methadone maintenance flexible doses be used for opioid maintenance treatment?

**Patient or population:** Opiate dependents.

**Setting:** Outpatients in United States, Australia, Austria, Switzerland, and UK.

**Viewpoint:** societal.

**Abbreviations:** RCT, randomized controlled trial; NA, not available; SD, standard deviation.

<sup>a</sup> Including dispensing fee.

<sup>b</sup> Includes staff time (i.e., face-to-face contact and preparation time), diagnostic procedures, and facility level (supplies, consumables, capital, equipment, ancillary support including administration, management, security, etc.).

<sup>c</sup> The study was conducted within the Australia health system, while the recommendation was global.

<sup>d</sup> This information was provided only by Harris et al. [10], and it was not considered because the risk of bias was considered too large.

M. Bruni et al. / Journal of Clinical Epidemiology ■ (2012)

(JCE 論文 #10 より)

■ Item #-27: アウトカムの相対的な重要度の再評価

Blog: 2012/08/17

どのアウトカムが重要なのかについては、システマティック・レビューやガイドライン作成に着手する前の、研究計画書作成の段階で決定する必要がある。その一方で、ガイドラインパネルは、研究計画書の作成後、または分析の実施後になってアウトカムの重要度(例: 深刻な有害作用)が明らかになる可能性についても留意し、このようなアウトカムをエビデンステーブルに含めるべく、適切な措置を講じるべきである。

GRADE JCE論文(\*)において、アウトカムの相対的重要度を考慮するための3ステップが発表されている。論文中の表1の内容が、GRADE教科書の表1.2-2(下記)である。

(\*)GRADE guidelines: 2. Framing the question and deciding on important outcomes

[Journal of Clinical Epidemiology](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3401441/) Volume 64, Issue 4, Pages 395-400, April 2011

表 1.2-2 アウトカムの相対的重要性を考慮するための3ステップ

ステップ	何をするのか?	なぜそうするのか?	どのように実施するのか?	根拠は?
1	エビデンスをレビューする前に、アウトカムを「重大」、「重要だが重大でない」、「重要ではない」に予備的に分類する。	エビデンスを検索、要約する際に最も重要であると考えられるアウトカムに注目し、かつ見解の不一致を解消、または明確にするため。	パネルや他の人々に、どのようなアウトカムが重要と考えられるかを特定してもらい、特定されたアウトカムの相対的重要性を判断してもらい、見解の不一致について議論してもらう。	これらの判断は、パネルの経験や、研究エビデンスに関する予備知識に基づいて下すことができる。パネル以外の患者に、アウトカムの相対的重要性を特定し、評価してもらうことの有用性も高い。
2	エビデンスをレビューした後、アウトカムの相対的重要性を再評価する。	当初は考慮されていなかったものの、エビデンスのレビューを通じて明らかになった重要なアウトカムが組み込まれること、かつ入手可能なエビデンスを考慮しながらアウトカムの相対的重要性を再検討することを確実にするため。	パネル(適切であればそれ以外の人々)に、第1ステップで組み込まれたアウトカムや、エビデンスのレビューによって明らかになった追加的アウトカムの相対的重要性的について再検討してもらう。	パネルやその他の情報提供者の経験、ならびに介入効果に関するシステマティック・レビュー。
3	介入の望ましい効果と望ましくない効果のバランスを判断する。	推奨を提示し、推奨度を決定するため。	パネル(適切であればそれ以外の人々)に、バランスシート(SoF テーブル)を用いることで、そして適切であれば、決断分析や経済分析を用いることで、望ましい効果と望ましくない効果のバランスを判断してもらう。	パネルやその他の情報提供者の経験、介入効果に関するシステマティック・レビュー、重要なアウトカムに対する患者の価値観に関するエビデンス(適切かつ入手可能であれば)、決断分析または経済分析(適切かつ入手可能であれば)。

■ Item #-28: その CQ の全体(overall)のエビデンスの質の判定

Blog: 2012/08/20

overall quality of evidence across outcomes :

推奨度を評価する際は、重大なアウトカム全般に関する全体的なエビデンスの質を判断する必要がある。GRADEでは、全体的なエビデンスの質を、「重大なアウトカムに関する複数のエビデンスから、最終的に“1つのエビデンスの質”として判断したもの」と定義している。

(GRADE教科書 p101)

アウトカム全般に関する全体的なエビデンスの質を判定する際の基準を表3.2.1-1に示す。実際には、複数の重大なアウトカムでエビデンスの質や効果の方向が異なる場合があり、その場合は、特に慎重にエビデンスの質を判断する。

表 3.2.1-1 アウトカム全般に関する全体的なエビデンスの質の判定基準

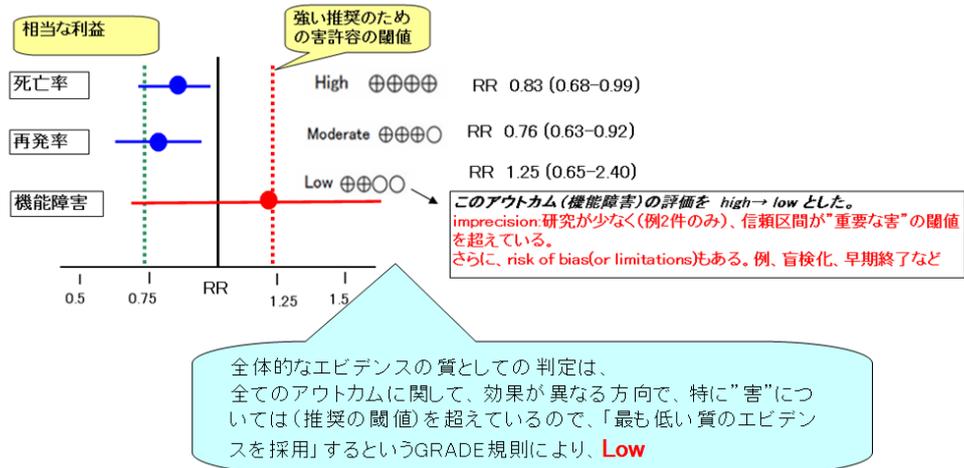
<ol style="list-style-type: none"><li>1. 重大 (critical) なアウトカムのみを検討する。</li><li>2. 重大なアウトカム間でエビデンスの質が異なり、かつ、<ul style="list-style-type: none"><li>• 各アウトカムが異なる方向（利益と害）を示している場合： 「重大なアウトカムに関するエビデンスの質の中で最も質が低いもの」を、全体的なエビデンスの質とする。</li><li>• すべてのアウトカムが同じ方向（利益、または害のいずれか一方）を示している場合： 「重大なアウトカムに関するエビデンスの質の中で最も質が高いもの」を、全体的なエビデンスの質とする。 ただし、利益と害のバランスが不確実である場合、重大なアウトカムのうち、「最もエビデンスの質が低いもの」とすべきである。</li></ul></li></ol>
---

**\*このルールはGRADEの特徴のひとつで、Minds基準とは正反対のものです。**

例：各アウトカムが異なる方向 (スライド # 91)

CPGIにおけるエビデンスの質評価例-1 (仮想)

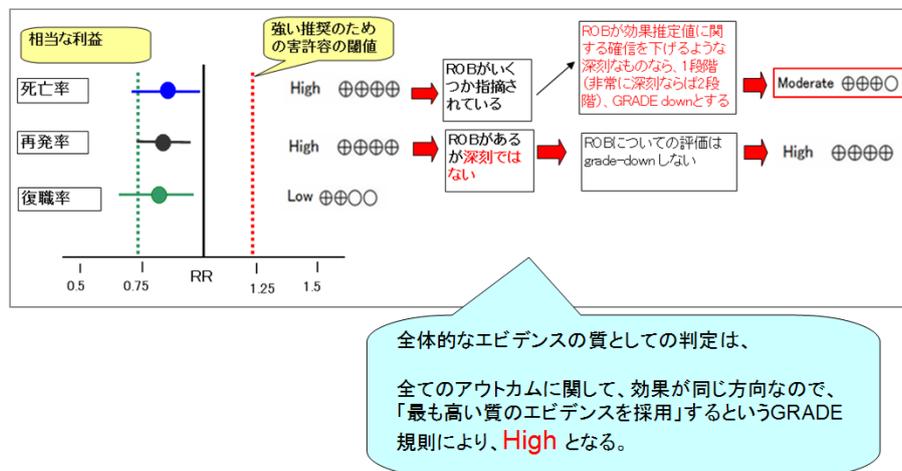
全般においてエビデンスの質が異なり、かつ各アウトカムが異なる方向を示す場合 (利益の方向と害の方向が異なる)



例：全てのアウトカムが同じ方向 (スライド # 93)

CPGIにおけるエビデンスの質評価例-3 (仮想)

各アウトカムについていずれも有効性を示す場合 (利益の方向と害の方向が同じ)

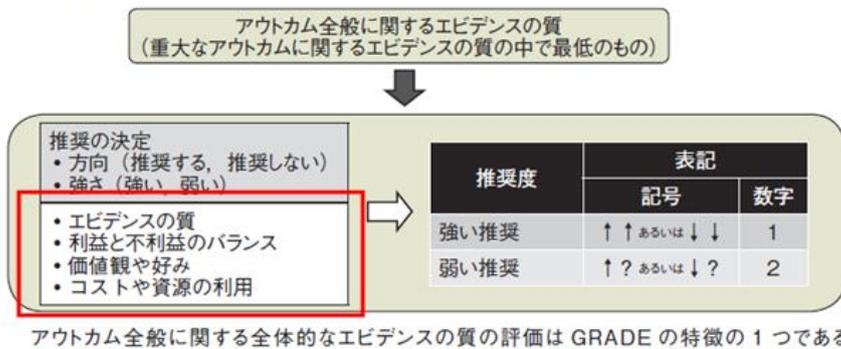


■ Item #-29: (推奨決定の4要因) 全体的なエビデンスの質

Blog: 2012/08/23

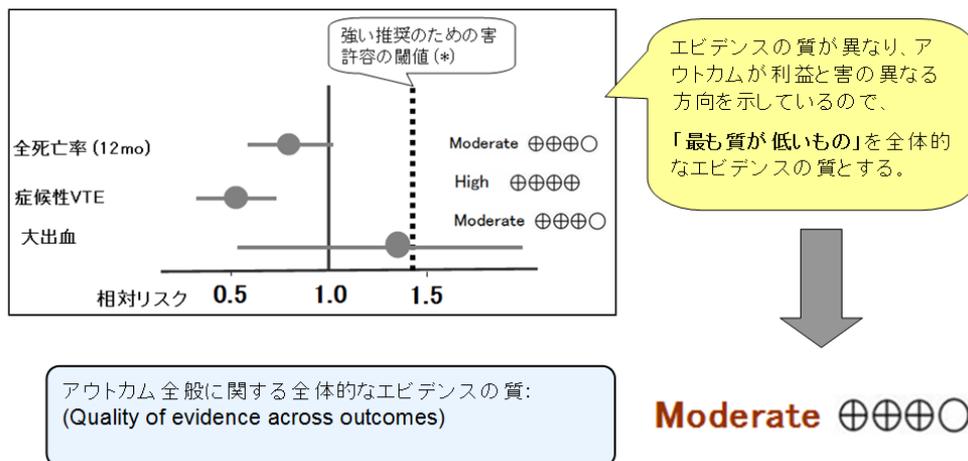
一般的に、質の高いエビデンスは強い推奨につながる可能性が高いが、推奨度はエビデンスの質だけで決定するものではなく、全体的なエビデンスの質、利益・不利益バランス、価値観や好み、コスト・医療資源利用の4要因を考慮して判定する。GRADEの推奨度の分類は、「強い」、「弱い」の2種類であるが、推奨の方向として、「実施する」、「実施しない」の2種類があるため、結果として推奨の表現は4通りとなる。推奨度は実際には連続的性質を有する概念であるが、GRADEではあえて「強い」、「弱い」の2値として表示する(図3.1)。

図 3-1 GRADE による推奨度の評価



利益と害のバランスをもとに治療を判断する際に、害に関するエビデンスの質が低い場合もある(例: 観察研究や症例集積研究においてサンプルサイズやイベント数が少ないとき)。そのような場合、GRADEでは、ある推奨の決定に不可欠なすべての重大なアウトカムのうち、エビデンスの質が最も低いものをもって、全体的なエビデンスの質と判定する。

・ シナリオの場合<sup>3</sup> (スライド #95)



<sup>3</sup> 閾値の設定を含めて、全て仮想の評価である。

## 附) エビデンスがない場合に、推奨は作成できないか？

ガイドラインパネルの決定方針で異なるだろう。

実際の報告例としては、以下のようなGRADEガイドラインがある。

(Canadian Task Force on Preventive Health Care, 2011)

### Screening for Breast Cancer

#### Summary of recommendations for clinicians and policy-makers

Recommendations are presented for the use of mammography, magnetic resonance imaging, breast self exam and clinical breast exam to screen for breast cancer. These recommendations apply only to women at average risk of breast cancer aged 40 to 74 years. They do not apply to women at higher risk due to personal history of breast cancer, history of breast cancer in first degree relative, known BRCA1/BRCA2 mutation, or prior chest wall radiation. No recommendations are made for women aged 75 and older, given the lack of data.

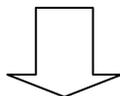
#### Mammography

- For women aged 40 – 49 we recommend not routinely screening with mammography.  
(Weak recommendation; moderate quality evidence)
- For women aged 50-69 years we recommend routinely screening with mammography every 2 to 3 years  
(Weak recommendation; moderate quality evidence)
- For women aged 70 - 74 we recommend routinely screening with mammography every 2 to 3 years.  
(Weak recommendation; low quality evidence).

#### Magnetic Resonance Imaging

- We recommend not routinely screening with magnetic resonance imaging  
(Weak recommendation; no evidence)

[http://canadiantaskforce.ca/recommendations/2011\\_01\\_eng.html](http://canadiantaskforce.ca/recommendations/2011_01_eng.html)



### Magnetic resonance imaging

*We recommend not routinely screening for breast cancer using MRI scans. (Weak recommendation; no evidence.)*

There are no data evaluating whether screening women at average risk of breast cancer using MRI scans reduces mortality as compared with mammography or no screening. Thus, screening women at average risk using MRI scans is not recommended.

■ Item #–30: 推奨決定(4要因)：望ましい効果と望ましくない効果のバランス

Blog: 2012/08/24

望ましい効果(利益)と望ましくない効果(害、負担、コストなど)のバランスで、望ましい効果が望ましくない効果を明らかに上回る、または明らかに下回る場合には、「強い推奨」と判定する。一方、両者が拮抗している場合や得失(トレードオフ)の関係の場合や、それぞれの効果あるいは両者のバランスのいずれかに不確実性がある場合は、「弱い推奨」の評価となる。

・シナリオ：スライド #100

NNTとNNHを計算してもバランスの評価は可能であるが、GRADEproを使って絶対効果を表示できる。GRADEproによるエビデンスプロファイル例を参照のこと。

[http://homepage3.nifty.com/aihara/EP\\_heparin\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/EP_heparin_aihara_20120426.html)

[http://homepage3.nifty.com/aihara/EP\\_heparin\\_accp\\_aihara\\_20120426.html](http://homepage3.nifty.com/aihara/EP_heparin_accp_aihara_20120426.html)

GRADE evidence profileにおいては、ベースラインリスクを3段階まで設定して、絶対効果を表示できる。

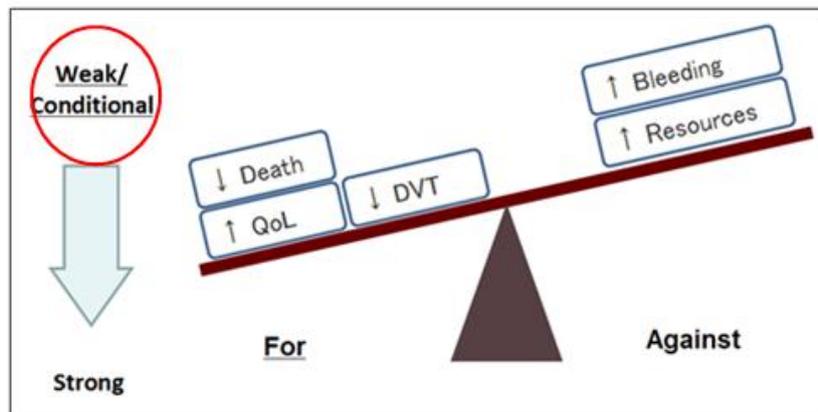
Outcome	No of participants (studies) Median follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with heparin versus no heparin (95% CI)
All-cause mortality	2530 (8 studies) 12 months	⊕⊕⊕⊕ MODERATE <sup>1,2,3</sup> due to inconsistency	RR 0.93 (0.85 to 1.02)	557 per 1000 645 per 1000 791 per 1000	39 fewer per 1000 (from 84 fewer to 11 more) 45 fewer per 1000 (from 97 fewer to 13 more) 55 fewer per 1000 (from 119 fewer to 16 more)
Symptomatic VTE	2264 (7 studies) 12 months	⊕⊕⊕⊕ HIGH <sup>1</sup>	RR 0.55 (0.37 to 0.82)	62 per 1000 29 per 1000	28 fewer per 1000 (from 11 fewer to 39 fewer) 13 fewer per 1000 (from 5 fewer to 18 fewer)
Major bleeding	2843 (9 studies) 12 months	⊕⊕⊕⊕ MODERATE <sup>1,4</sup> due to imprecision	RR 1.3 (0.59 to 2.88)	19 per 1000 7 per 1000	6 more per 1000 (from 8 fewer to 35 more) 2 more per 1000 (from 3 fewer to 13 more)
Minor bleeding	2345 (7 studies) 12 weeks	⊕⊕⊕⊕ MODERATE <sup>1,4</sup>	RR 1.05 (0.75 to 1.46)	51 per 1000	3 more per 1000 (from 13 fewer to 23 more) 1 more per 1000 (from 7 fewer to 12 more)

対照群リスクを、全体では56%、moderate riskとして65%、high riskとして79%とした。

1000人あたりリスク差を表示することで、利益と害の比較が容易である。

・スライド #97

本例では、弱い(条件付き)推奨と評価した。



■ Item #-31: 推奨決定(4要因): 価値観と好み

Blog: 2012/08/29

価値観や好みの不確実性、または患者間におけるばらつきが存在する場合は、推奨の強さが低くなる。

どのような管理戦略にも必ず長所と短所があり、患者は常にそれらの間の得失を評価しないとけない。そのため、ガイドラインパネルがある特定の利益、リスク、不便さに対してどのような価値を置くのかによって、推奨や推奨の強さが決まる。

各アウトカムに対する価値観や好みの問題について調べた研究がごく限られているため、価値観や好みに関しては常に大きな不確実性があるとする意見もあるかもしれない。一方、価値観や好みに関する系統的な研究もあり、ガイドラインパネルメンバーも患者との関わりを通して新たな知見を得てきた。(GRADEハンドブック:[価値観や好みにおける確信について](#))

シナリオ: スライド #106

患者にとって重大なアウトカムのエビデンスの質は「中 (moderate)」である。

推奨: 「抗凝固療法の適応がない癌患者に対して、非経口的抗凝固療法の実施を提案する」

**Weak (↑?) / moderate quality of evidence**

価値観と好み: 「この推奨作成においては、全死亡率や症候性VTEの減少に大きな価値をおき、ヘパリンによる出血に比較的低い価値を置いた。また、外来通院や非経口的な投与の負担にも低い価値を置いた」

■ Item #-32: 推奨決定(4要因)：資源の利用(コスト)

Blog: 2012/08/30

一般に、医療に関するコストや資源の利用に関しては、形式的経済学モデル(formal economic model)を用いて1単位あたりの便益(利益)を得るために必要なコストとして表現する。(GRADEハンドブック：[形式的な経済モデルの有用性について](#))

どのようなコスト算出にどのモデルを用いるかの例をあげると、たとえば脳梗塞を1件予防するためにかかるコストを算出するには費用効果分析(cost-effective analysis)を、非金銭的価値としての質調整生存年(Quality-Adjusted Life Years: QALY)あたりのコストを算出するためには費用効用分析(cost-utility analysis)を、金銭的価値で示される単位あたりの利益を得るために必要なコストを算出するためには費用便益分析(cost-benefit analysis)を用いる。

このような経済モデルを用いた研究結果は意思決定の際の参考にはなるものの、残念ながら、その多くが限界やバイアスを含む可能性があり、モデルの前提とされた特定の状況に限定された結果であることが多い。そのため、GRADEでは、通常のアウトカムについてのエビデンスプロファイルには費用効果分析または費用効用分析を含めないことを推奨している<sup>4</sup>。

(GRADE教科書 p107)

シナリオ：スライド #108

抗凝固療法による全死亡率(12か月)の効果、NNT=18 (10～63)の結果から、単純な薬剤費用としての必要コストを計算すると、低分子ヘパリン (dalteparin 2500単位) 平均20w(\*)では、1人約7万円である。

したがって、1件の全死亡(12ヶ月)を減らすための利用コストは、70万円 (NNT 10)～450万円 (NNT 63)となる。

(\*) ヘパリン投与期間の20wは、レビューにおいて記載されている論文データの平均値である。

<sup>4</sup> 経済評価(economic evaluation) には、費用帰結研究(cost-consequences study), 費用最小化研究(cost-minimization study), 費用効果分析(cost-effectiveness analysis), 費用効用分析(cost-utility analysis), 費用便益分析(cost-benefit analysis) などがある。

■ Item #-33: 推奨度評価の合意形成(必要なら GRADEgrid)

Blog: 2012/08/31

GRADEで採用するシステムはRAND法<sup>5</sup>に準じたもので、重要・重大なアウトカムの選出やエビデンスの質評価にもこの統計手法の使用が可能である。

参考：GRADE教科書 (p128-131)

付録4.1 「パネルがGRADEを適用する際のコツ」

付録4.2 「推奨度評価の合意形成(GRADE grid)」

シナリオ：スライド #115(ここでは、RAND/UCLA法を利用)

## 修正版Delphi法： RAND/UCLA

合意のための討議は、エビデンスの質や推奨の評価だけではなく、アウトカムの重要性評価にも必要ですが、グループの討議内容を集約するために使われるのが 修正版デルファイとしてのRand法である。

評価点の中央値(median)、評価点の三分位の頻度(1～3、4～6、7～9を選択した評価者数)、見解不一致指数(パーセンタイル範囲 / 対称を補正したパーセンタイル範囲)を算出する。<sup>1)</sup>  
見解不一致指数は、パネリスト数の大小にかかわらず算出でき、ばらつきを示す指標としては、中央値からの平均絶対偏差よりも効率的である。指数が1を超える場合は、見解の不一致を意味する。

### 1<sup>st</sup> Round rating

appropriate: 評価点の中央値が7点以上、かつ不一致なし  
uncertain: 評価点の中央値が4～6点、あるいは不一致あり  
inappropriate: 評価点の中央値が3点以下、かつ不一致なし



2nd Round rating

GRADEにおける推奨は、強さを2種類(強い、弱い/条件付き)、方向を2種類(推奨する、推奨しない)の組み合わせにより、4種類のいずれかに決定する。判定不能<sup>2)</sup>の決定をしないという原則で、「討議によって合意にいたらない場合は、GRADE gridを適用する」という方針である。

<sup>5</sup> Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998; 2: i-iv, 1-88. PMID: 9561895

シナリオ：スライド #113～ 117

GRADEproを使った推奨の決定と GRADEgridの仮想データ例：

[http://www.grade-jpn.com/grade\\_grid\\_20120420.html](http://www.grade-jpn.com/grade_grid_20120420.html)

メンバーの70%(14/20)が“介入の実施”を支持し、推奨の強さ weakが50%(10/20)。

GRADE	Strong	Weak	No strength	Weak	Strong
<b>Assessors view of the balance of desirable and undesirable consequences of the intervention</b>	Desirable consequences clearly outweigh undesirable consequences	Desirable consequences probably outweigh undesirable consequences	Consequences equally balanced or uncertain	Undesirable consequences probably outweigh desirable consequences	Undesirable consequences clearly outweigh desirable consequences
<b>Recommendation</b>	We recommend to do something	We suggest to do something	No specific recommendation	We recommend to not do something	We recommend to not do something
We suggest parenteral heparin for patients with cancer who have no other therapeutic or prophylactic indication.	4	10	3	2	1

Strength of a recommendation	WEAK
------------------------------	------

Gridは(あらゆるガイドライングループにとって)、GRADEアプローチを使って合意を形成し、エビデンスの解釈を取り巻く不確実性のパターンをより明確に理解する上で有益である。

CPGに示す、GRADEの推奨の強さは、strong, weakの2種類であり、判定不能は採択しない。

■ Item #34: 診療ガイドライン作成

Blog: 2012/09/04

ガイドラインパネルは、推奨の草案がエビデンスプロファイルのデータに沿っていることを確認し、様式の統一性、推奨表現の一貫性を確保する。

例1: スライド # 111

推奨作成 癌患者における非経口的抗凝固療法(ヘパリン)

<p><b>疑問:</b> 非経口的抗凝固療法を癌患者の生存延長のために使用すべきか <b>患者:</b> 癌患者(抗凝固療法の適応のない、肺癌、大腸癌、前立腺癌など)。 <b>介入:</b> 非経口ヘパリン <b>比較:</b> ヘパリン投与なし <b>アウトカム:</b> 全死亡率、大出血、小出血、症候性DVT、QOL、血小板減少。</p> <p><b>エビデンスの要約:</b>(以下は重大なアウトカムのみについて記載する) 8件のRCTのシステマティックレビュー(患者2530人)では、全患者における死亡率(12ヶ月)には統計的に有意な差はなかった(RR 0.93, 95%CI 0.85~1.02; NNT 18, 10~63)。サブグループ解析では小細胞肺癌患者において有意な差があった(RR 0.86, 95%CI 0.75~0.98; NNT 9, 95%CI 5~40)。 7~9件のRCTのメタアナリシス(各々、患者 2264人、2843人)では、統計的に有意な症候性VTEの減少(RR 0.55, 95%CI 0.37~0.82; NNT 31, 95%CI 20~61)と大出血の増加(RR 1.30, 95%CI 0.59~2.88; NNH 2532, 95%CI 96~104)が認められた。</p> <p><b>エビデンスの質:</b> 全死亡率に関するエビデンスの質は、非一貫性(inconsistency: I<sup>2</sup>=41%の中等度の異質性あり)。サブ解析結果は小細胞肺癌患者における生存率改善を示唆)と不精確さ(imprecision)とをあわせて、グレードを1段階下げて「中」とした。症候性VTEに関するエビデンスの質は、グレードを下げる要因に深刻な問題はなく、「高」であった。大出血に関するエビデンスの質は、不精確さ(利益と害の両方の可能性を含む)のために、グレードを「高」から1段階下げて「中」とした。</p> <p><b>最良の推定値:</b> 小細胞肺癌患者の死亡率減少、症候性VTEの減少が示されたが、全患者における死亡率は差がなかった。</p> <p><b>利益と害、負担、コストの判定:</b> 利用できる情報では死亡率減少の利益が示唆され、NNTsとNNHを考えると、ある患者(例えば、小細胞肺癌)あるいは出血リスクが少ない患者はヘパリンによる抗凝固療法を受け入れるであろう。悪性疾患ということを考慮すると、ほとんどの患者が外来通院や非経口的な薬剤投与を受け入れるだろう。コストはそれほど高くない。</p> <p><b>推奨のグレーディング:</b> 患者にとって重要なアウトカムのエビデンスの質は「中」である。 推奨表現:「凝固療法の適応がない癌患者に対して、非経口的抗凝固療法の実施を提案する(条件付きの推奨/「中」のエビデンスの質)」。 価値観や好み: この推奨作成においては、全死亡率や症候性VTEの減少に大きな価値をおき、ヘパリンによる出血に比較的に低い価値を置いた。また、外来通院や非経口的な投与の負担にも低い価値を置いた。</p>
--

例2: [http://www.cadth.ca/media/pdf/compus\\_IA\\_OT\\_rec\\_report.pdf](http://www.cadth.ca/media/pdf/compus_IA_OT_rec_report.pdf)

2.2.1 Type 1 diabetes mellitus in children

CERC recommends that either insulin lispro or insulin aspart be used in **children with type 1 diabetes using CSII** if treatment with a rapid-acting insulin analogue is chosen.

Vote Results (Number of CERC Members in Favour/Number Participating in Meeting)	10/11
CERC Rating of Overall Quality of Clinical Evidence	Moderate
Strength of Recommendation	Weak
<p><b>Underlying values and preferences:</b></p> <ul style="list-style-type: none"> <li>Primary consideration: The desirability of allowing patients and clinicians to choose between agents similar in cost and effectiveness.</li> </ul>	
<p><b>Context:</b></p> <ul style="list-style-type: none"> <li>Overall quality of evidence was low.</li> <li>Studies primarily reported results regarding surrogate outcomes. There were no data on the effects of insulin aspart versus insulin lispro on clinically important long-term outcomes.</li> <li>There was no significant difference in A1C.</li> <li>The rate of overall hypoglycemia was slightly reduced in the insulin aspart arm versus insulin lispro. No other differences in hypoglycemia were observed.</li> <li>For patients using CSII, the insulin pumps frequently need to be removed (i.e., due to line blockage or sporting activities); hence, the same insulin for MDI and CSII should be used.</li> </ul>	

A1C=hemoglobin A1C; CERC=COMPUS Expert Review Committee; CSII=continuous subcutaneous insulin infusion; MDI=multiple daily injection.

各推奨事項の一般形式として決まったものはないが、参考としてACCPより発表されている様式に関する記述を下表に示す。

ACCP Evidence-Based Guideline Development: A Successful and Transparent Approach Addressing Conflict of Interest, Funding, and Patient-Centered Recommendations, Michael H. Baumann, Sandra Zelman Lewis, David Gutterman CHEST 2007; 132:1015–1024

ACCPガイドライン： 推奨事項の様式および形式
推奨の最初の部分には、適切な患者集団を示す。制約要因があれば、それも記載する。
推奨では、たとえばサービス、検査、治療、または医療措置を提案する。
推奨の文言に用いる表現としては、事実について述べるような表現ではなく、サージ巢に関する確定的な提案を行うような表現を使うこと。一覧表または表の形式は使用しないこと。事実の供述およびその根拠については、推奨の直前の本文に記載する。利益が明らかにリスクおよび負担を上回る場合は、「推奨する (recommend)」などといった強い表現を使用する。利益と害のバランスがそれほど明確でない場合は、「提案する (suggest)」、「考慮する (consider)」などといった表現を使用し、適切な状況および患者集団に関する適切な説明を添えること。
推奨には引用を含めない。引用は、推奨の直前に記載される考察の適切な場所に含める。
推奨のグレーディングは、推奨の最後に提示される。
グレーディングのためのシステムは、ACCPのHSP評価システム(GRADEシステム)に従う。
推奨は太字で記載する。

## ガイドライン執筆パネルの最終会議

### *The Final EBG Writing Panel Conference*

ACCP Evidence-Based Guideline Development (CHEST 2007; 132:1015-1024)の、“The Final EBG Writing Panel Conference” 項を翻訳したものである。EBG: evidence-based clinical practice guideline

-----  
エビデンステーブルが完成し、執筆パネルによる章の初期ドラフトの執筆、レビュー、修正が完了すると、執筆パネルの最終対面会議が開催され、パネルによる最終版の受理が確定される。この会議では、未解決の重要な問題および紛争を明らかにし、パネルメンバー全員で、これについて議論し、必要であれば投票により問題の解決が図られる。

議論および検討の対象となるのは、推奨事項の文言、リスクと利益のバランス、(パネルが決定していない場合は)エビデンスの質の解釈、推奨の合意の程度(例、執筆パネルの80%以上が合意した場合<sup>6</sup>、その推奨は承認される)である。

ただし、議長の裁量により、(1)圧倒的多数(supermajority >80%)の合意が得られるまで議論を継続する場合、(2)合意不十分を理由に推奨を削除する場合、(3)推奨は削除されないが、推奨の直前の考察に少数意見を含める場合がある。

-----  
<sup>6</sup> 各ガイドラインパネルによって判定基準は異なる。  
例、敗血症救命キャンペーンガイドライン (GRADE 教科書 p131)。

## ■ Item #-35: レビューと外部評価

### EBG Review

Blog: 2012/09/05

ACCP Evidence-Based Guideline Development (CHEST 2007; 132:1015-1024)の、“EBG Review” 項を参考にしたものである。EBG: evidence-based clinical practice guideline

-----  
最終会議の議事録に基づいて修正が加えられたガイドライン原稿は、レビューおよび改訂のためにガイドラインパネルの複数のグループに提出される。担当グループは、プロセス、一貫性、推奨および評価が適切かどうか、および内容についてレビューする。執筆グループは、レビューの批判を十分に吟味した上で、ガイドライン原稿を評議会に提出し、承認を得る。

評議会のメンバーは、構造化されたレビューフォームをしようすることにより、徹底的なレビューの実施を確実にしている。このレビューフォームの一部は、[AGREE](#) (Appraisal of Guidelines Research and Evaluation) ツール<sup>7</sup>をベースとする([ACCPレビューフォーム](#); 次頁)。

評議会の承認が得られると、原稿は出版のための審査および独立した外部組織によるレビューが実施される。

ACCPのエビデンスに基づく診療ガイドラインのレビュープロセスでは、徹底性と迅速性の両立が図られ、なおかつ最大限の質および透明性が確保されている。このプロセスの迅速性は、エビデンスのレビューの終了と同時に推奨の適時性が減じられていくという観点から重視される。

-----  
すなわち、ガイドラインの公開前には、エビデンスの検索と要約の方法、推奨の作成方法などの質を、なんらかの方法で評価すべきである。ガイドラインの質のチェックには、いくつかの手法が開発報告されており、WHOのGWGチェックリスト([附録C「WHO 治療ガイドライン用チェックリスト」154 頁](#))、Appraisal of Guidelines for Research & Evaluation (AGREE)、Conference on Guideline Standardization (COGS)などがある。(GRADE教科書 p126)

<sup>7</sup> 現在は、AGREE-II が利用されている。

ACCP ガイドラインレビューフォーム
<p><u>診療ガイドラインの概要 (Overall outline)</u></p> <ol style="list-style-type: none"> <li>1. 診療ガイドラインにおいて、ガイドラインの目標、および具体的な臨床課題が明確に述べられている。</li> <li>2. 組み入れ／除外基準について明確な説明がされている。</li> <li>3. ガイドラインの対象ユーザーについて適切な説明がある。</li> </ol>
<p><u>方法論 (Methodology)</u></p> <ol style="list-style-type: none"> <li>4. 説明のされた方法論を使って、診療ガイドラインが作成されている。</li> <li>5. 文献のシステマティック・レビューが実施されている。</li> <li>6. エビデンスの質は、正式なシステムを用いて評価付けされている。</li> <li>7. エビデンス、および利益と害の評価に基づいた推奨である。</li> </ol>
<p><u>推奨の提示 (Presentation)</u></p> <ol style="list-style-type: none"> <li>8. 推奨内容が具体的でわかりやすい。</li> <li>9. 患者集団について具体的に説明されている。</li> <li>10. 診療ガイドラインの中で、主要な推奨事項が明確にわかるようになっている。</li> <li>11. 推奨事項の要約(臨床アルゴリズムを含む場合あり)が提供されている。</li> </ol>
<p><u>診療への適用可能性 (Applicability)</u></p> <ol style="list-style-type: none"> <li>12. 診療ガイドラインにおいて、推奨を実行するための戦略が示されているか。</li> <li>13. 診療の中でどのようにガイドラインを活用するのかについての具体的情報が提示されているか。</li> </ol>
<p><u>信頼性 (Accountability)</u></p> <ol style="list-style-type: none"> <li>14. 診療ガイドラインの中で資金提供者が開示されているか。</li> <li>15. ガイドライン作成組織の全メンバーにより、利害衝突の可能性が開示され、明確な説明があったか。</li> </ol>
<p><u>関連性およびわかりやすさ (Relevance and readability)</u></p> <ol style="list-style-type: none"> <li>16. 臨床的に関連性があるか。</li> <li>17. わかりやすいか。</li> <li>18. 意味がわかるか。</li> <li>19. その考察はエビデンスから(エビデンスが存在する場合)導き出されているか。</li> <li>20. 推奨はその考察から導き出されているか。</li> </ol>
<p><u>レビューのステータス (Review status)</u></p> <ol style="list-style-type: none"> <li>21. この資料は、ACCPのガイドラインとして承認を得られるか。             <ul style="list-style-type: none"> <li>・そのままの状態承認</li> <li>・提案内容を検討中のため承認を保留(追加コメントが添付のグリッドにて提供される場合あり)</li> <li>・承認しない(グリッドに示されるコメントを参照)</li> </ul> </li> </ol>

(AGREEより改変) <http://journal.publications.chestnet.org/article.aspx?articleid=1085372>

## ■ Item #-36: 配布および普及

### Dissemination and Marketing

Blog: 2012/09/06

以下は、ACCP Evidence-Based Guideline Development (CHEST 2007; 132:1015-1024)の、“Dissemination and Marketing” 項を参考にしたものである。EBG: evidence-based clinical practice guideline

ACCP Evidence-Based Guideline Development: A Successful and Transparent Approach Addressing Conflict of Interest, Funding, and Patient-Centered Recommendations, Michael H. Baumann, Sandra Zelman Lewis, David Gutterman CHEST 2007; 132:1015-1024

-----  
ACCPのほとんどのEBGにて、「臨床リソースツール」が作成される。このツールは、診療ガイドラインに基づく、印刷およびCD-ROMの組み合わせによる実行ツールキットである。標準コンポーネントとしては、以下が含まれる。

- ◆ 臨床医のためのクイック参考ガイド (*Quick Reference Guide for Clinicians*)
- ◆ 患者教育資料
- ◆ プレゼンテーション用スライド: 医療分野の専門家および非専門家を対象  
(さらに医師用の指示書などの追加資料を追加することも可能である)

“臨床医のためのクイック参考ガイド”には、プリント、電子フォーマット、または携帯情報端末(PDAなど)にダウンロード可能な形式で、全ての臨床アルゴリズム、および評価付けされた主な推奨事項が盛り込まれている。

-----  
例 スマートガイドライン<sup>8</sup>

<http://www.skillupjapan.co.jp/news/2012/0709.asp>

<sup>8</sup>スキルアップジャパンと東京慈恵会医科大学は 2012年7月9日、パソコンやモバイル端末から診療ガイドラインを検索できる Web サービス「スマートガイドライン」を共同開発し、サービスの提供を開始した。

## ■ Item #-37: ガイドラインの管理

### Guideline Maintenance

Blog: 2012/09/07

以下は、ACCP Evidence-Based Guideline Development (CHEST 2007; 132:1015-1024)の、“Guideline Maintenance” 項を翻訳したものである。

EBG: evidence-based clinical practice guideline

HSP: Health and Science Policy Committee

ACCP Evidence-Based Guideline Development: A Successful and Transparent Approach Addressing Conflict of Interest, Funding, and Patient-Centered Recommendations, Michael H. Baumann, Sandra Zelman Lewis, David Gutterman CHEST 2007; 132:1015-1024

-----  
エビデンスに基づく診療ガイドライン(EBG)は、HSP (and by appropriate NetWorks\* and in consultation with the original EBG Executive Committee)によって毎年レビューされ、推奨事項の内容が現状に沿っているか、または中間研究によって改訂を許可するのに十分な情報が提供されているかどうかの判断が行われる。最初にガイドラインが出版されてから1年を起点として、各ガイドラインには、年に一度、次に示すステータスカテゴリの一つが付与される。

1. ガイドラインの内容は現状に即しており、年に一度レビューする。
2. 有用かもしれない新たなエビデンスが存在する。ただし、この新たなデータは推奨内容を変更するのに十分ではないと判断されることから、修正は認可されない。
3. 本ガイドラインの項/章の修正が認可されるような新たなエビデンスが存在する。ACCPはこのEBGの修正を検討中である。
4. 新たなエビデンスが相当量存在し、現行のガイドラインが現状に即さないため、廃止する。

-----  
\*NetWorks are the ACCP interdisciplinary, expertise-focused interest groups that provide college members the opportunity for personal and professional alliances with the ACCP.

■ Item #-38: 質の改善および実行

*Quality Improvement and Implementation*

Blog: 2012/09/10

以下は、ACCP Evidence-Based Guideline Development (CHEST 2007; 132:1015-1024)の、“Quality Improvement and Implementation” 項を参考にしたものである。

ACCP Evidence-Based Guideline Development: A Successful and Transparent Approach Addressing Conflict of Interest, Funding, and Patient-Centered Recommendations, Michael H. Baumann, Sandra Zelman Lewis, David Gutterman CHEST 2007; 132:1015-1024

-----  
診療ガイドラインは、質が確保され、その改善や実施、および責任の範囲と所在が明確に記載されているべきである。たとえば、ACCP では、診療ガイドラインの推奨内容の改善や実施は、適切なパフォーマンス指標(\*)にしたがって、実行可能性(feasibility)、有用性(usability)、科学的 중요度(scientific importance)、実用性(practicality)、適用性(applicability) を確保しなければならない、としている。(GRADE教科書 p127)

パフォーマンス評価指標は、ACCP メンバーおよびその患者にとって実用的かつ適切なものである必要がある。患者のケアを改善するためのパフォーマンス評価指標のベースとなる推奨事項と、パフォーマンス評価指標のベースとすべきでない推奨事項(多くはエビデンスの質が低い場合が多い)も特定する。これに加え、介入の一部が実行不可能または利用不可能な場合もある。パフォーマンス評価指標は、現在、米国の医療分野における各種セクターにより企画されているペイフォーパフォーマンス指標(価値に基づくパフォーマンス(value-based performance)戦略とも呼ばれる)の中に盛り込んでもよい。ペイフォーパフォーマンス戦略が成功を収め、患者および医療提供者の双方にとって公正なものとなるかどうかは、実行されたパフォーマンス評価指標の質と密接に関係する。

-----  
(\*)パフォーマンス評価指標の基準: [National Quality Forum](#)

■ 最後に:

Blog: 2012/09/12

GRADE を採用する場合の注意

GRADE を使ったシステマティック・レビューや診療ガイドラインを作成する場合、次の3つの注意が必要である。

1. GRADE を改変しない。ぜひとも GRADE を使用したいという熱意から、GRADE を変更してしまう組織がある。GRADE プロセスの変更はエビデンスやガイドライン利用者の混乱を招くかもしれない。また、このような変更は臨床医、政策決定者、患者が使用できる単一のシステムを提供するという目的に反する。以上の理由から、GRADE ワーキンググループは、このような変更を行わない事を推奨する (GRADE 教科書 p190 脚注参照)。
2. GRADE の“長所”と“限界”を認識する  
(GRADE 教科書 p8-9、GRADE の使いかたスライド #122～124 参照)。
3. GRADE ワーキンググループが作成した、「GRADE システムを利用したと言えるための基準」を参考にする。(スライド #125)

**GRADE システムを利用したと言えるための基準<sup>9</sup>**

1. 「エビデンスの質 quality of evidence」は、GRADE Working Group が採用する 2 つの定義(ガイドラインまたはシステマティック・レビューのいずれか)により、一貫して定義すべきである。
2. エビデンスの質を評価するための各 GRADE 基準 (risk of bias/研究の限界、直接性、結果の一貫性、精確さ、出版バイアス、効果の大きさ、用量反応勾配、説明のつかない交絡やバイアス「相反バイアス (antagonistic bias)」の影響) を、用語の違いがあるにしても、明確に記述すべきである。
3. 各重要なアウトカムの全体的なエビデンスの質 (overall quality of evidence) を、4 段階 (例: 「高 high」、「中 moderate」、「低 low」、「非常に低 very low」)、あるいは正当性が認められるならば、3 段階 (例: 「高 high」、「中 moderate」、「低 low (「低」および「非常に低」を含む) )」にて、GRADE Working Group が採用する定義に合致した各段階の定義に基づき、評価ならびに等級付けすべきである。
4. エビデンスの質ならびに推奨度は、エビデンスの要約 (ナラティブまたは表形式) に基づき判断すべきである。理想的には、GRADE Working Group が提唱する完全版エビデンスプロファイルを使用すべきである。また、エビデンスプロファイルはシステマティック・レビューをベースとすべきである。少なくとも、評価されたエビデンス、ならびにそのエビデンスの同定や評価に使用した手法を明確に記述すべきである。特に、グレードアップやグレードダウンの理由についてはわかりやすく説明すべきである。
5. 推奨の強さ (strength of a recommendation) を評価するための各 GRADE 基準について明確に考慮し (望ましい帰結と望ましくない帰結のバランス、エビデンスの質、価値観と好み、資源の利用状況)、一般のアプローチを報告すべきである (例: コストを考慮したかどうかや、その方法、ならびに誰の価値観と好みを前提としたのかなど)。
6. 管理選択肢に対する肯定的または否定的な推奨の強さは、2 段階 (「弱い weak/条件付き conditional」、「強い strong」) で示すべきである。また、各段階の定義は、GRADE Working Group が採用

<sup>9</sup> 今後出版予定の JCE 論文をもとに改訂が必要となる可能性があります。

する定義に合致すべきである。推奨度を示すのに「弱い／条件付き」、「強い」以外の用語を使用する場合でも、その解釈や内容は、GRADE Working Group の定義に合致すべきである。

7. 理想的には、推奨度の判断をわかりやすく報告すべきである。

GRADE JCE シリーズとは

Blog: 2012/09/13

2008年のBMJシリーズ6篇の後、システマティック・レビューや医療技術を評価する著者、ガイドラインパネリストやガイドラインパネルをサポートする方法論学者などにGRADEアウトプット(エビデンスプロファイル、SoF テーブル、グレード付けされた推奨など)を作成するための詳細な指針を提供することを目的とした新シリーズとして、2011年よりGRADE-JCE論文が連載発表されています(Box 1<sup>10</sup>、2012年9月現在、第12編まで発表)。

<p><b>Box 1 GRADE <i>Journal of Clinical Epidemiology</i> series—list of articles</b></p> <p><b>Introductory articles</b></p> <ol style="list-style-type: none"><li>1. Introduction and summary of findings tables</li><li>2. Framing the question and deciding on the importance of outcomes</li><li>3. Rating the quality of evidence—introduction</li></ol> <p><b>Rating confidence in estimates of effect (quality of evidence)</b></p> <ol style="list-style-type: none"><li>4. Rating the quality of evidence—risk of bias</li><li>5. Rating quality of evidence—publication bias</li><li>6. Rating the quality of evidence—imprecision (random error)</li><li>7. Rating the quality of evidence—inconsistency</li><li>8. Rating the quality of evidence—indirectness</li><li>9. Rating up the quality of evidence</li><li>10. Rating the quality of evidence for resource use</li></ol> <p><b>Summarizing the evidence</b></p> <ol style="list-style-type: none"><li>11. Summarizing the quality of evidence for individual outcomes and across outcomes</li><li>12. Preparing summary of findings tables—binary outcomes</li><li>13. Preparing summary of findings tables—continuous outcomes</li></ol> <p><b>Diagnostic tests</b></p> <ol style="list-style-type: none"><li>14. Applying GRADE to diagnostic tests—summarizing the evidence</li></ol>	<p><b>Making recommendations</b></p> <ol style="list-style-type: none"><li>15. Going from evidence to recommendations—the meaning of strong and weak recommendations</li><li>16. Going from evidence to recommendations—determinants of a recommendation’s direction and strength</li><li>17. Going from evidence to recommendations—resource use</li><li>18. Going from evidence to recommendations—diagnostic tests</li></ol> <p><b>Observational studies</b></p> <ol style="list-style-type: none"><li>19. Special challenges in using observational studies</li><li>20. GRADE for public health and health policymaking</li></ol> <p><b>Concluding articles</b></p> <ol style="list-style-type: none"><li>21. Group processes, variations of GRADE, and further developments of GRADE part 1</li><li>22. Group processes, variations of GRADE, and further developments of GRADE part 2</li></ol> <p><i>Abbreviation:</i> GRADE, Grading of Recommendations Assessment, Development, and Evaluation.</p>
--	--

GRADE guidelines—an introduction to the 10th–13th articles in the series.  
Guyatt GH, Oxman AD, Schünemann HJ. J Clin Epidemiol. 2012 Sep 8

<sup>10</sup> 2012年9月のJCE論文(J Clin Epidemiol. 2012 Sep 8)発表により、G.H. Guyatt et al. / Journal of Clinical Epidemiology 64 (2011) 380–382 の table1 を Box1 の内容に変更した(2012/0919)。

GRADE JCE シリーズ紹介論文の序文の一部を翻訳紹介し、本解説集 (GRADE ガイドライン作成の流れ:解説) の最後としたい。

GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology (Journal of Clinical Epidemiology 64 (2011) 380-382)

#### Introduction

\*\*\*\*\*

世界各国の医療専門家、研究者、ガイドライン作成者により結成される Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group は2000年、エビデンスの質を評価し、診療ガイドラインにおける推奨の強さを判断するための最適なシステムの開発に向けて一丸となり取り組みを開始した。GRADE Working Groupは現在200名を超えるメンバーを抱え、10年間の取り組みを経た現在もなおGRADEの方法を改善し、拡大するために会合している。これまでに開催された25回に及ぶ1～2日間の会合と数え切れないほどのメールのやり取りは、診療や医療判断のための研究エビデンスを解釈する方法を開発・改善し、そのエビデンスを臨床医、患者、政策決定者に最適な形で提示するためのラボのような存在となっている。

～(略)～

今のところ、このシリーズの最終編では観察研究にみられる特別な課題に関する論文を1つ、グループプロセス、GRADEのバリエーション、今後のGRADEの展開予想についてのGRADE Working Groupの見解を示す論文を2つ紹介する予定である。今後の展開に関しては、このシリーズに掲載される方法論がGRADE適用に関する不変の決定的指針として存続するものと期待すべきではなく、またそうなることはありえない。このシリーズは数多くの方法論的問題にアプローチするためのアドバイスを提供する。中でも革新的アプローチの一例としては、代理エンドポイントへの対処方法、不精確さに起因する限界に関する判断基準、サブグループ解析の信頼性評価基準、診断検査に関するエビデンスの質の判断、連続変数の効果の大きさに関する要約があげられる。

革新的アプローチには固有の不安定性があり、すべてではないにしてもその一部のアプローチについては改善が必要不可欠である。また、将来的には革新的方法のみでなくすでに確立されている概念についても方法論的進展や改善があるだろう。われわれはシステム適用のための一連の不変の基準を提示することはできないが、GRADEの利用者にはどうか落胆しないでいただきたい。このシリーズで紹介する広範かつ包括的な基礎知識は現時点または今後システマティック・レビューやガイドライン、医療技術評価にGRADEを適用する際に多いに役立つものであることに違いはないからだ。本誌編集者らは、本シリーズがゆくゆくは臨床疫学の発展の1つの節目として認められることになることを自負している。

\*\*\*\*\*

[目次へ](#)