

GRADE

GRADE システムの使い方

エビデンスから推奨へ

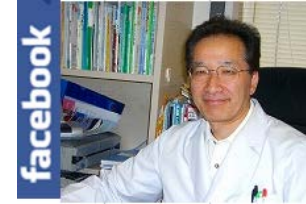
How to use GRADE system

Going from evidence to recommendations

相原守夫 Morio Aihara, M.D.
(member of the GRADE Working Group)

相原内科医院 青森県弘前市青山3-8-2
TEL: 0172-33-0055
ezzy01757@nifty.ne.jp

April 26, 2012



Morio aihara, M.D.
Hirosaki, Japan

<http://www.facebook.com/morio.aihara>



本資料の作成および、改訂情報

今後の内容改訂については、このセクションで重要な変更のまとめを示す。今後は、本資料の新しいバージョンをダウンロードするたびにこのセクションを参照していただきたい。

スライドの追加により番号が変わる可能性があるため、[ハーパーリンク](#)を作成している。

- 2012.09.18: GRADEガイドライン作成の流れ: 解説リンクをスライド#4に追加
- 2012.08.23: GRADEの略: Grades → Grading
- 2012.06.12: スライド#[105](#)(シナリオに関する推奨のまとめ)を追加、[#110](#)(修正版Delphi, RAND法)を追加
- 2012.06.08: スライド#[69](#), [#70](#) (対応リスクの計算), [#71](#) (Evidence profile表示の改変)を新規追加
スライド#[85](#), [#86](#), [#87](#) (CPGにおけるエビデンスの質評価例を3件)追加
- 2012.05.31: スライド#[120](#)新規追加、スライド [#99](#)の脚注に、”Decision table”、feasibility(実行可能性)を記載
- 2012.05.23: スライド# [67](#), [#88](#): heparin(-)群の分母 1060→1066に変更(n=2530)
- 2012.05.02: web公開 (how_to_grade_aihara_20120426.ppt)
- 2012.04.25: Gordon Guyatt, Holger Schunemannら、海外のGRADEメンバーにppt (ver.20120425) 送付
- 2012.04.06: 「GRADEシステムの使い方」 ver. 20120406 作成開始

I have had the support and encouragement of Holger J Schunemann, Gordon Guyatt, Yngve Falck-Ytter, Jan L Brozek and Elie Akl. I would like to appreciate their support very much.

[GRADEハンドブック翻訳版\(ハンドブック情報\)](#)に記載してあるが、エビデンスの質と推奨の強さをグレーディングするためのGRADEハンドブックを複製および翻訳する場合は、編集者からの許可を得ること(Holger J Schunemann, Jan Brozek, Andrew Oxman)。

GRADEハンドブック (ver. 3.2)翻訳版(*)は、非営利目的、かつ引用サイトを明らかにしていただければ、自由に利用してかまいません。

GRADE: Grading of Recommendations Assessment, Development and Evaluation



<http://www.gradeworkinggroup.org/index.htm>

online learning modules: <http://cebgrade.mcmaster.ca/>

GRADEに関する日本語の情報源

相原内科医院 webサイトにて公開しています。

<http://www.grade-jpn.com/>

- 診療ガイドラインのためのGRADEシステム (相原 他、凸版メディア出版、2010)

<http://homepage3.nifty.com/aihara/grade-cover.html>

- GRADEハンドブック (version 3.2 翻訳版、Free)

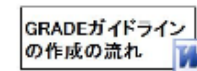
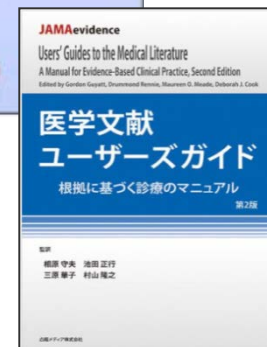
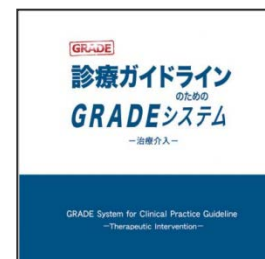
<http://homepage3.nifty.com/aihara/GRADEproHelp.html>

- 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
(JAMA Users' Guides翻訳 : 相原 他、凸版メディア出版、2010)

http://homepage3.nifty.com/aihara/jama_evidence.html

- GRADEガイドライン作成の流れ: 解説 http://www.grade-jpn.com/grade_gl_flow_aihara.doc

本解説(DOC) は、http://www.grade-jpn.com/grade_gl_flow_20120717_2.pdf の項目に対応した内容です。



GRADE Working Group

GRADE:

Grading of Recommendations Assessment, Development and Evaluation

- GRADE ワーキンググループは、医療で用いられる既存のグレーディング・システムの問題点を改善したいと願う人々により、非公式の共同グループとして2000年に発足した。当グループの目的は、エビデンスの質と推奨の強さをグレーディングするための、一般的かつわかりやすい方法を開発することである。
- 2000年から、ガイドライン開発を代表する人たちが国際的に共同して作成および評価を行っている。

内容

Part-1 エビデンスの質

Part-2 エビデンスから推奨へ

Part-3 GRADEの長所・限界と使用基準

注意:

本資料では、エビデンスの質や推奨の強さをReview manager 5.1とGRADE profiler (version 3.6)を使って評価しているが、GRADEシステムを解説するために個人的にデータを作成したものである。

利用しているCDSR例の詳細は、オリジナル論文を読んでいただきたい。

Part 1

エビデンスの質

Quality of Evidence

エビデンスの確信性

There always is evidence

“When there is a question there is evidence”

エビデンスは常に存在する

「疑問のあるところにはエビデンスが存在する」

エビデンスは、極端に弱いものかもしれない(1人の臨床医による非系統的な観察や、非直接的に関連するにすぎない作用機序を示す生理学的研究など)、必ず存在する。

Better research ⇒ greater confidence in the evidence and decisions

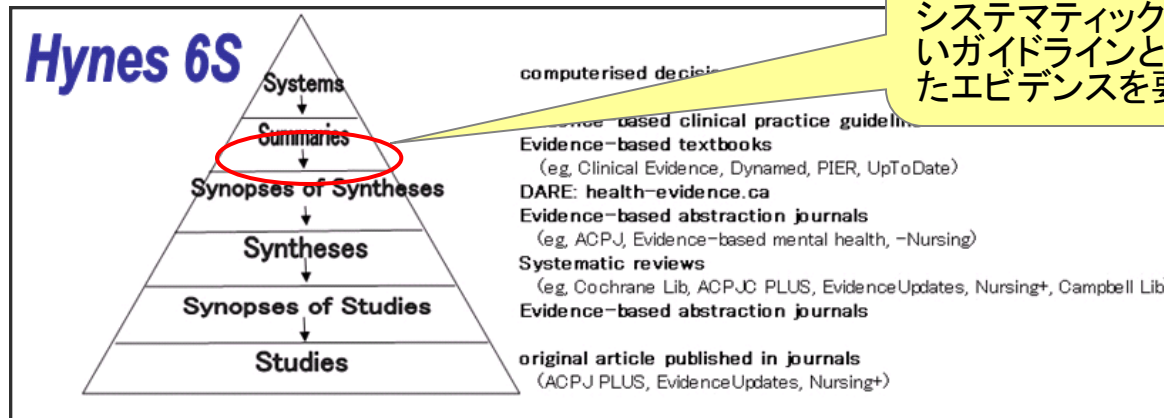
より良い研究 → エビデンスや判断の信頼性が高い。

Evidence alone is never sufficient to make a clinical decision

臨床決断を下すのにはエビデンスだけでは決して十分ではない。

エビデンスとは？

臨床疑問があるところには必ずエビデンスがある。



例： サマリー (Summaries)

システマティック・レビューや質の高いガイドラインとして、1次研究から得たエビデンスを要約し提示。

http://homepage3.nifty.com/aihara/sakusaku/2_1.html

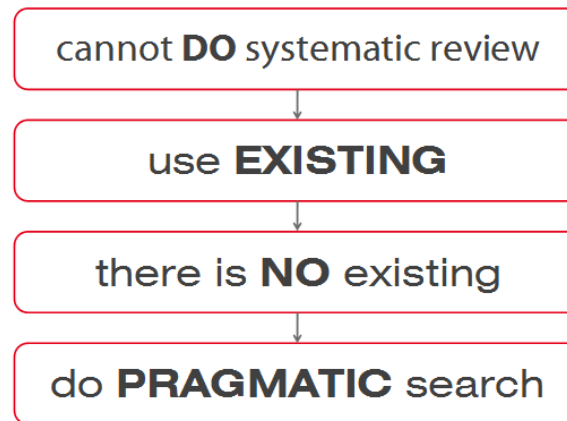
エビデンスの検索にはさまざまな方法がある。現実的には、Hynesの6sの上位にある情報源を使うことも労力を少なくして高質なエビデンスを抽出する方法の一つである。

エビデンスの質の定義は、システマティック・レビューと診療ガイドラインで異なる。

推奨のためにシステマティックレビューを利用する

GRADEシステムを利用した診療ガイドラインの作成には、原則としてシステマティックレビューの総体エビデンス (body of evidence) を用いる。

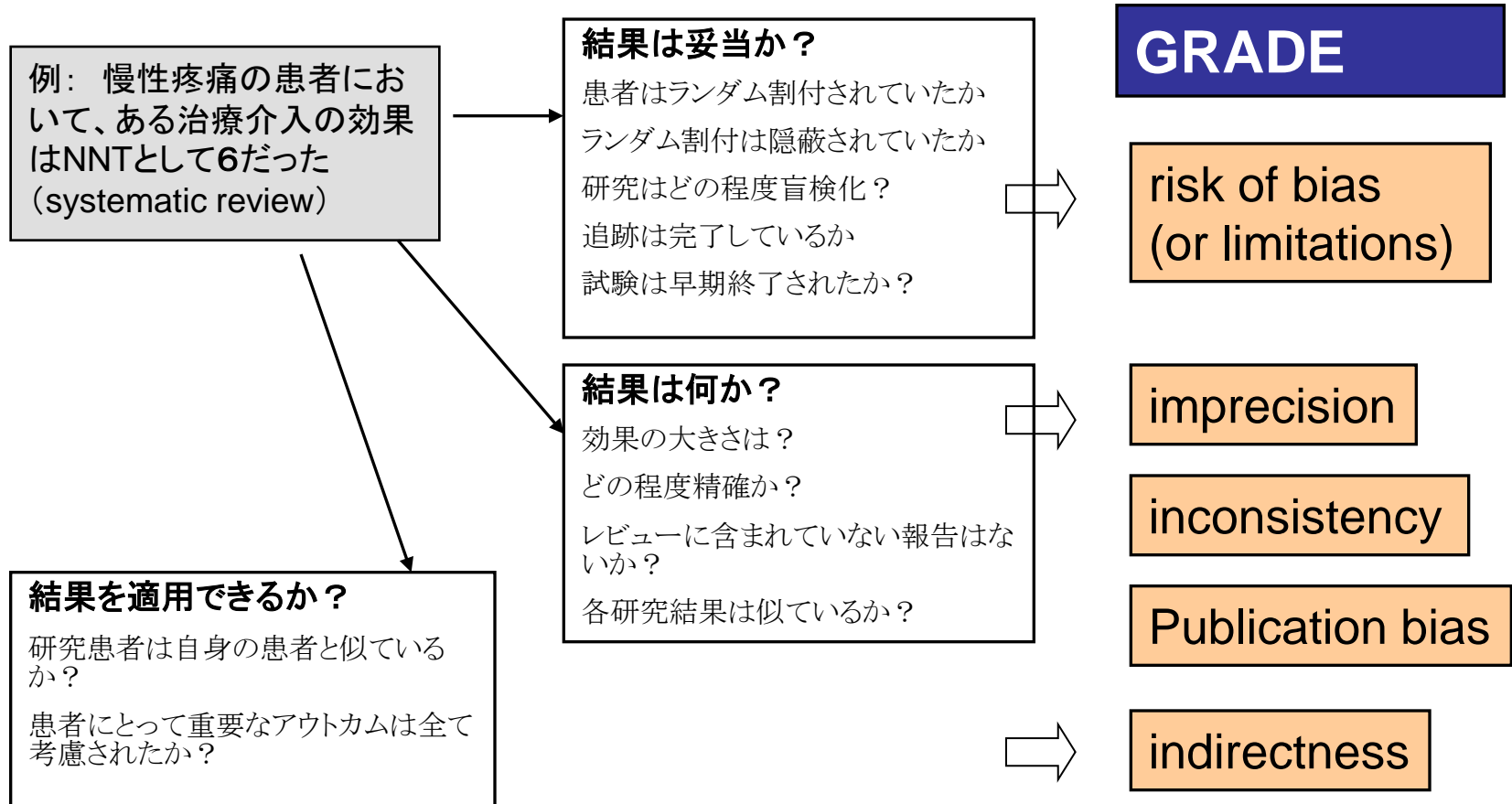
質の高い既存のシステマティックレビューを利用することも可能であるが、ガイドラインパネルが選択した全ての懸案のアウトカムに関するエビデンスが1つのシステマティックレビューで提供されていることは稀であり、通常は1つの body of evidence の作成に複数のシステマティックレビューが必要となるため、多大な時間と労力が必要となる。



なぜ、システマティックレビューや診療ガイドラインに、GRADEを使うのか？



GRADEを使う論拠



エビデンスの質(定義)

■ システマティック・レビュー

The quality of evidence reflects the extent to which we are confident that an estimate of the effect is correct.

「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」

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 総体エビデンス(body of evidence)の質のグレーディングを行う(overall quality for **each** outcome, across studies)。

■ 診療ガイドライン

The quality of evidence reflects the extent to which our confidence in an estimate of the treatment effect is adequate to support particular recommendation

「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」

診療ガイドラインでは、“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を1つだけ決定する(overall quality **across** outcomes)。

エビデンスの質の分類

効果推定値に対する“確信の程度”は連続的なものであるが、“エビデンスの質”を4つのカテゴリーに分類する。

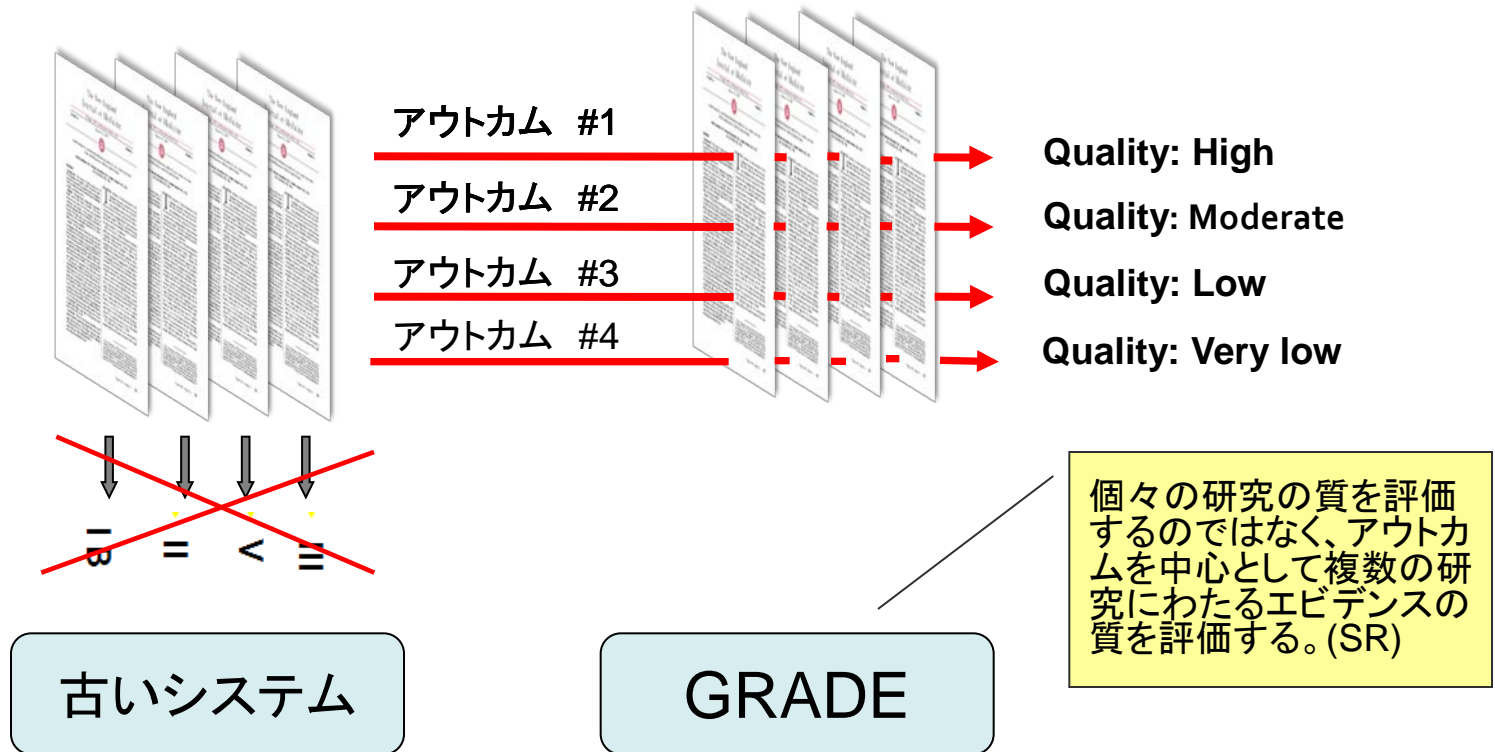
[J clinical Epidemiology 64\(4\): 401-406, 2011](#)

エビデンスの質	最近の定義 (JCE)*	以前の定義 (BMJ)
High ⊕⊕⊕⊕ / A	真の効果が効果推定値に近いという確信がある(*)。	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)
Moderate ⊕⊕⊕○ / B	効果推定値に対し、中等度の確信がある。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。	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)
Low ⊕⊕○○ / C	効果推定値に対する確信には限界がある。真の効果は効果推定値とは大きく異なるかもしれない。	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)
Very Low ⊕○○○ / D	効果推定値に対しほとんど確信がもてない。真の効果は、効果推定値とは大きく異なるものと考えられる。	あらゆる効果推定値が不確実である(very uncertain)

最近の定義も、以前の定義も基本的な内容は同じである。

GRADE はアウトカム中心である

GRADE is outcome-centric



システマティック・レビュー

疑問の定式化
アウトカムの選択
重要度の評価

複数の研究ア
ウトカムの統合

エビデンス・プロ
ファイル作成
(GRADEprofiler)

各アウトカムに
関するエビデン
スの質を評価

RCTは「高」から、観察
研究は「低」から開始

P
I
C
O

アウトカム 重大
アウトカム 重大
アウトカム 重要
アウトカム

重要でない



Quality assessment	1	2	3	4	5	6	7	8	9	10
1. Study limitations										
2. Imprecision										
3. Indirectness										
4. Inconsistency										
5. Publication bias										
6. Large effect										
7. Dose-response										
8. Confounding by indication										
9. Confounding by other factors										
10. Confounding by bias										

高
中
低
非常に低

グレードを
下げる

1. 研究の限界
2. 非一貫性
3. 非直接性
4. 不精確さ
5. 出版バイアス

グレードを
上げる

1. 大きな効果
2. 用量反応
3. 交絡因子

各アウトカムに関する
効果推定値と結果要約
(Summary of Findings)

ガイドライン作成

推奨の作成:

- 推奨または 推奨しない(方向)
- 強いまたは 弱い(推奨度)

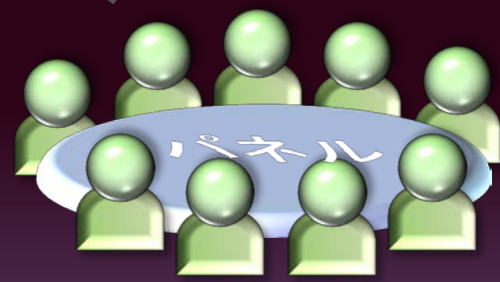
以下を考慮して判断:

- エビデンスの質
- 利益と害のバランス
- 価値観と好み



必要に応じて:

- 資源の利用(コスト)
- を考慮し改訂する

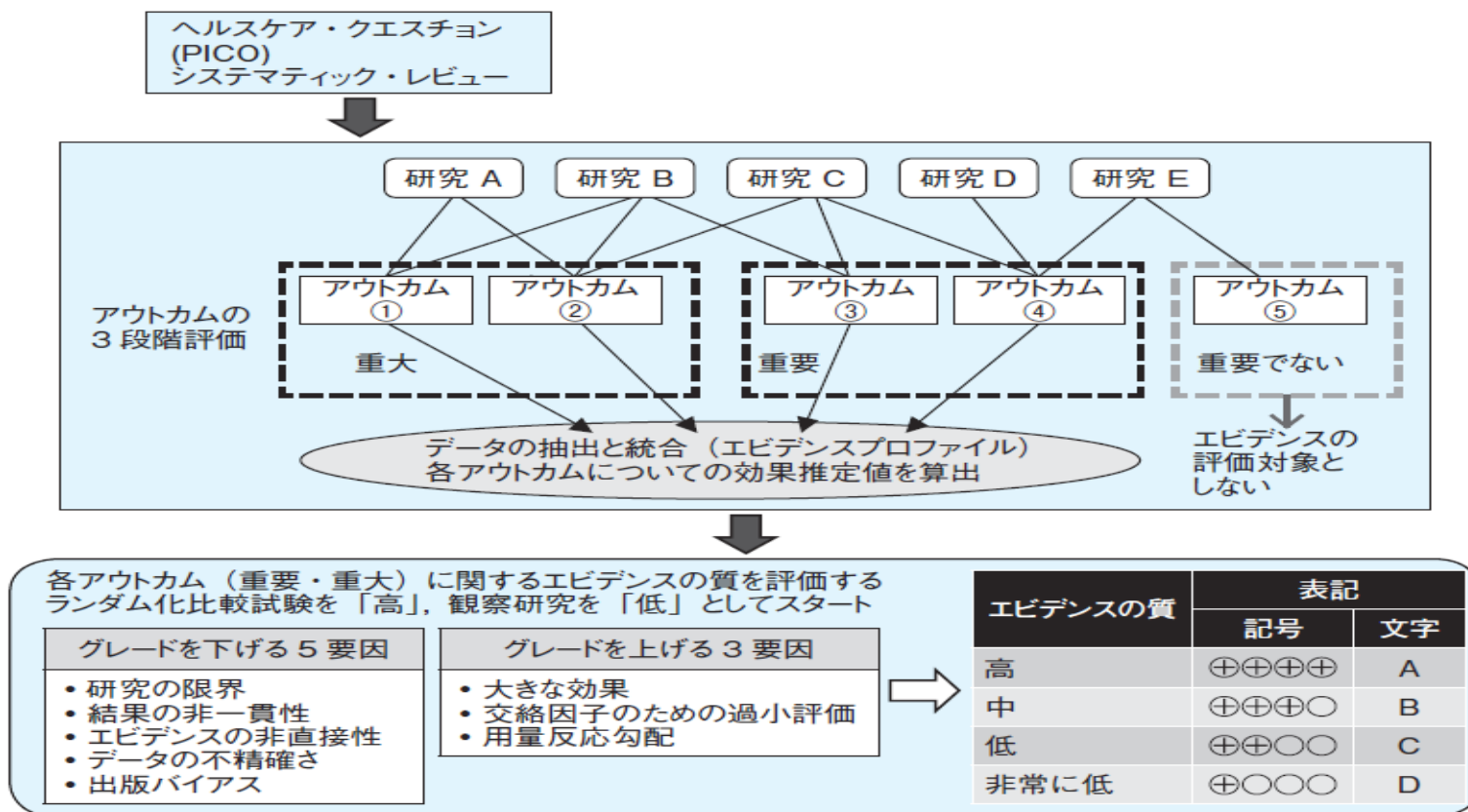


全体的なエビデンスの質
アウトカム全般に関する評価は、
重大なアウトカムに関する
エビデンスの質の中で最低のもの

- "...を使用する(強い推奨)"
- "...を使用する(弱い推奨)"
- "...を使用しない(強い推奨)"
- "...を使用しない(弱い推奨)"

システマティックレビューのエビデンスの質

図 1-1 GRADE によるエビデンスの質の評価フロー



「診療ガイドラインのためのGRADEシステム」相原守夫ら：凸版メディア社出版、2010年

臨床シナリオ -①

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療している。

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。

臨床疑問 (CQ)-1

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

PICOは？

- **Patient:** 癌患者 (肺癌、大腸癌、前立腺癌など)
- **Intervention:** parenteral anticoagulation (heparin)
- **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

患者にとって重要なアウトカムとは

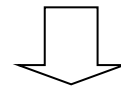
A patient important outcome is defined as follows: an outcome for which one would answer by yes the following question:

“if the patient knew that this outcome was the only thing to change with treatment, would the patient consider receiving this treatment if associated with side effects or cost?”

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。」

具体的にはどのようなアウトカムがあるだろうか？



アウトカムの重要度の階層

レビューは、アウトカムの階層構造を参考にして重要度を評価する。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。

I. Mortality

1. all cause mortality
2. disease specific mortality

II. Morbidity

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

III. Symptoms/Quality of life/Functional status (e.g. failure to become pregnant, successful nursing/breastfeeding, depression)

IV. Surrogate outcomes (e.g. diagnosis of TB, viral load, physical activity, weight loss, post operative atrial fibrillation, cognitive function, recurrent (malignant) polyps, adherence to medication)

アウトカムの重要度 (GRADE) -①

評価 (Rating)	重要度 (Importance)
9	重大
8	
7	
6	重要
5	
4	
3	重要でない
2	
1	

GRADEでは、アウトカムの重要度の評価に 9-point scaleを用いるよう提案している。

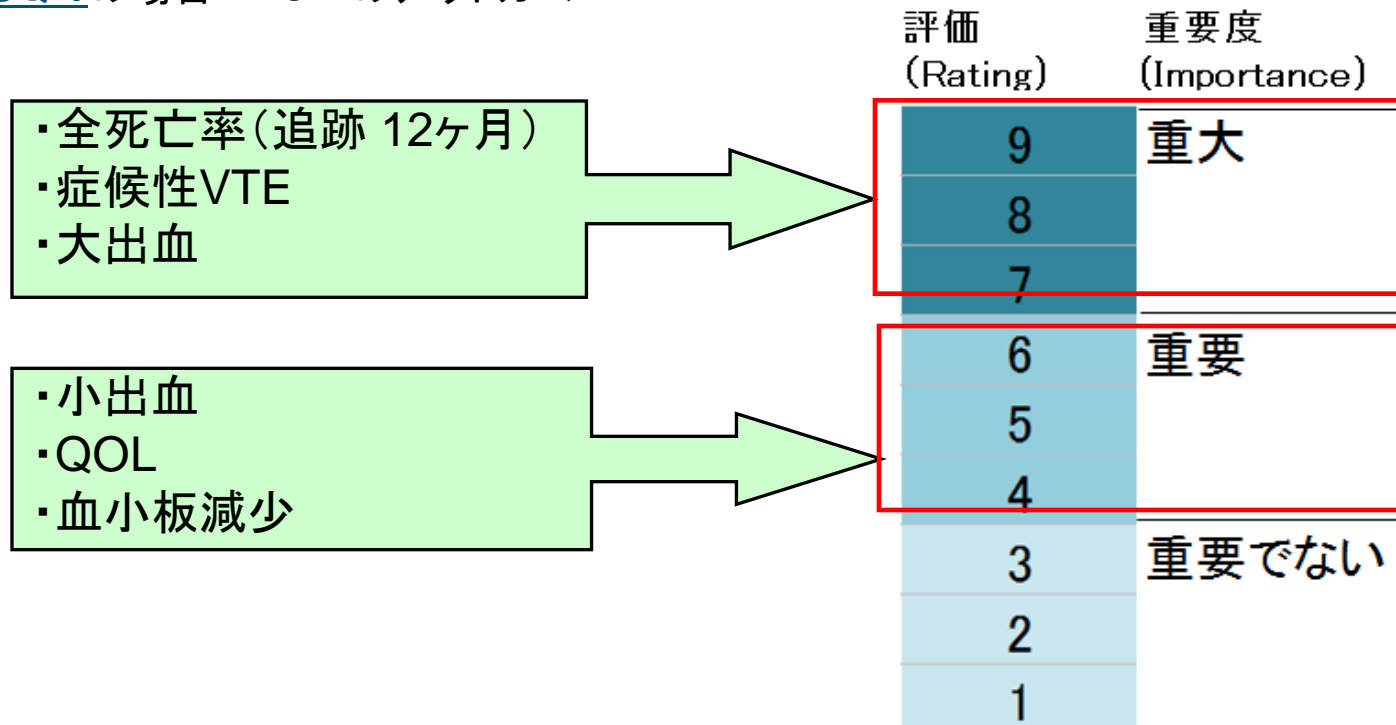
- ・重大： 7～9点
- ・重要だが重大でない： 4～6点
- ・重要でない： 1～3点

GRADEハンドブック(ver. 3.2) Section 6「アウトカムの重要性」

<http://homepage3.nifty.com/aihara/GRADEproHelp/Outcome%20importance.htm>

アウトカムの重要度 (GRADE) -②

CQ-1の場合・・・6つのアウトカム



アウトカムの重要性の評価は、エビデンス検索後に再評価が必要である

通常、どのアウトカムが重要なのかについては、システムティック・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性を再評価すべきである。

アウトカムの重要性は、レビューの作成者やガイドラインパネルの合意で決める。具体的には、まずパネルがアウトカム候補を提案・作成し、次に、どれが重要なアウトカムかを匿名で評価する。評価の平均または中央値をもとにRAND法などの統計学的手法を使用し、「重要」、「重大」なアウトカムを特定する。

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	何をやるのか?	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ムティック・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ムティック・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）。

エビデンスの検索とデータの抽出・統合

GRADE では、医療にかかわる推奨は、コクランライブラリ(Cochrane Library) のようなシステマティック・レビューに基づいたエビデンスを利用して作成することを推奨している。

システマティックレビューで収集されたエビデンスは、エビデンスの質評価と結果要約 (Summary of findings: 介入群と対照群における効果サイズ、相対効果と絶対効果の大きさ) を含むエビデンステーブルを作成するために使われる。つまり、エビデンスの質評価と結果要約を含むGRADE [エビデンスプロファイル](#) は、推奨作成者が自身の判断の基礎とする透明性の高いエビデンスの要約である。

GRADEハンドブック(ver. 3.2) Section 6「関連するエビデンスの要約」

<http://homepage3.nifty.com/aihara/GRADEproHelp/Systematic%20review.htm>

本CQ-1については、以下のコクランレビューを使う。



Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation (Review)

Akl EA, Gunukula S, Barba M, Yosuco 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>

エビデンスの検索式

MEDLINE

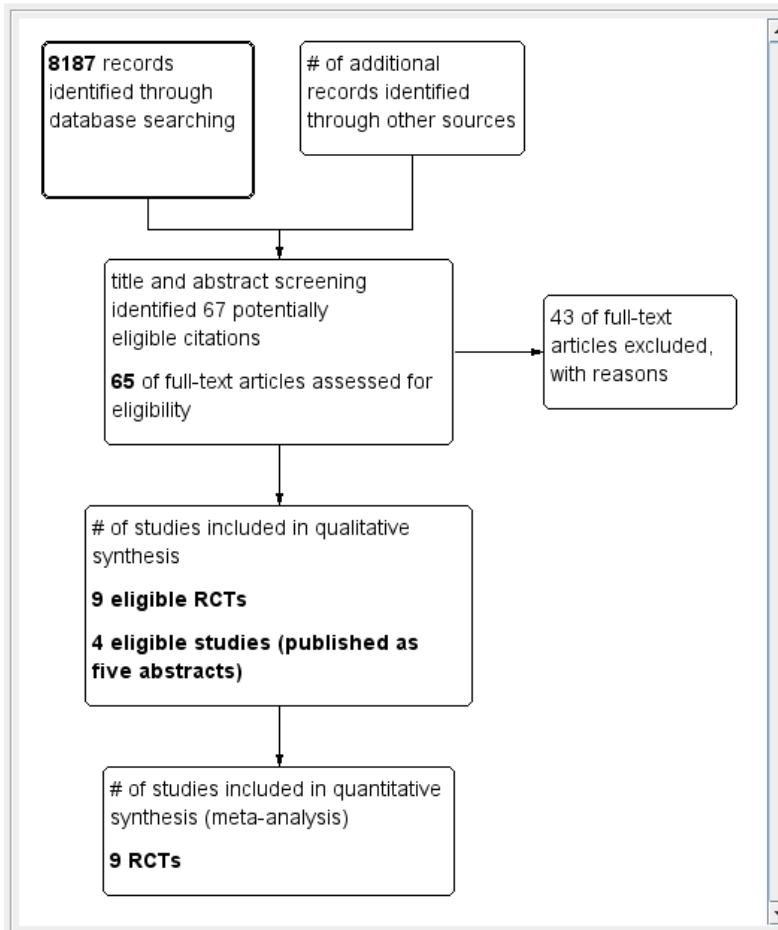
#1 Heparin/
#2 Heparin.tw
#3 Heparin, Low-Molecular-Weight/
#4 (LMWH OR low molecular weight heparin OR nadroparin OR
fraxiparin OR enoxaparin OR clexane OR lovenox OR dalteparin OR
fragmin OR ardeparin OR normiflo OR tinzaparin OR logiparin OR
innohep OR certoparin OR sandoparin OR reviparin OR divarin 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

CENTRAL (The Cochrane Library, latest issue)

#1 heparin OR low molecular weight heparin OR LMWH OR
lowmolecular-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
#2 Coumarins OR Warfarin OR coumadin OR acenocumarol OR
phenprocumon OR 4-hydroxicoumarins OR oral anticoagulant OR
vitamin K antagonist OR VKA
#3 fondaparinux OR Arixtra
#4 ximelagatran OR Exanta
#5 Pradaxa or Dabigatran or rivaroxaban or Xarelto or apixaban
#6 1 OR 2 OR 3 OR 4 OR 5
#7 malignan\$ OR neoplasm\$ OR cancer OR carcinoma\$ OR
#8 6 AND 7

検索において、“言語”による制限は
していない

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	Agnelli 2009
2	Altinbas 2004
3	Kakkar 2004
4	Klerk 2005
5	Lebeau 1994
6	Pelzer 2009
7	Perry 2010
8	Sideras 2006
9	Weber 2008

これらの各研究のアウトカムはどのようなものだったのだろうか？

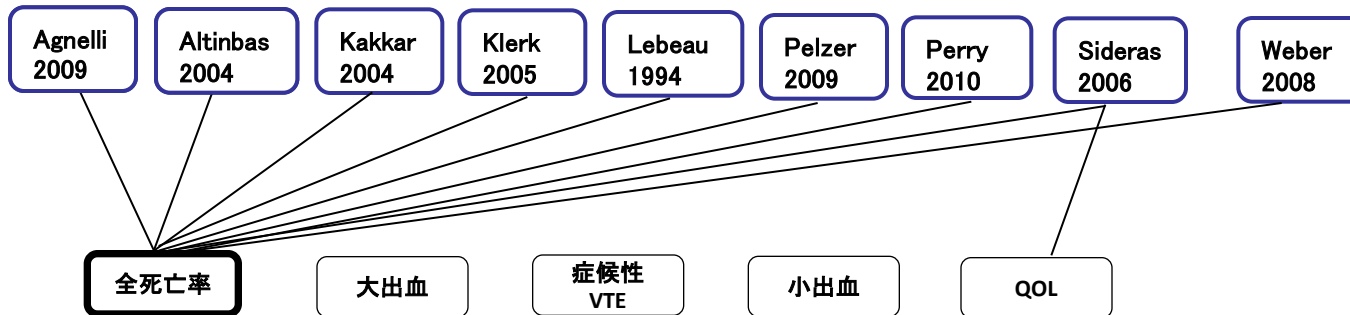


	研究	アウトカム 1	アウトカム 2	アウトカム 3	アウトカム 4	アウトカム 5
1	Agnelli 2009	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
2	Altinbas 2004	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
3	Kakkar 2004	all-cause mortality	major bleeding	minor bleeding	symptomatic VTE	-
4	Klerk 2005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

(-: 報告なし)

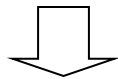
研究によっては、アウトカムの報告がないものがある

我々が知りたいのはアウトカムごとに研究データを統合した結果である。



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、総体エビデンス(body of evidence)と呼び、このエビデンスの質を評価する。

エビデンスの質の評価 (GRADE)

「システマティックレビューのエビデンスの質評価」と同じ内容

研究デザイン (Study design)	エビデンスの質 (Initial quality of a body of evidence)	グレードを下げる 5要因 (Lower if)	グレードを上げる 3要因 (Higher if)	エビデンスの質 (Quality of a body of evidence)
RCT	High 	<ul style="list-style-type: none"> ・バイアスのリスク (Risk of bias) -1 Serious -2 Very serious ・非一貫性 (inconsistency) -1 Serious -2 Very serious ・非直接性 (indirectness) -1 Serious -2 Very serious ・不精確さ (imprecision) -1 Serious -2 Very serious ・出版バイアス (publication bias) -1 Likely -2 Very likely 	<ul style="list-style-type: none"> ・効果が大きい (large effect) +1 Large effect RR > 2 あるいは < 0.5 +2 Very large RR > 5 あるいは < 0.2 ・用量-反応勾配 (dose-dependent gradient) +1 あり ・全ての交絡因子 (plausible confounder) +1 提示された効果を減弱させている +1 効果が観察されなかった場合に、効果を増加させる方向に働いている 	
観察研究	Low 			

エビデンスの質	表記	
	記号	文字
高	⊕⊕⊕⊕	A
中	⊕⊕⊕○	B
低	⊕⊕○○	C
非常に低	⊕○○○	D

Confidence (確信性)の決定要因

RCTs ⊕⊕⊕⊕

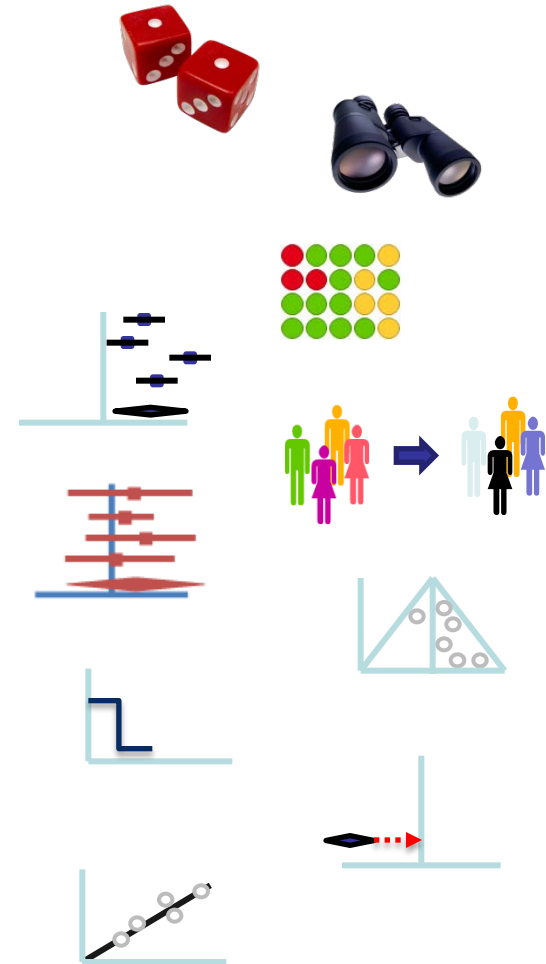
observational studies ⊕⊕○○

5 factors that can lower quality

1. limitations in detailed study design and execution (*risk of bias criteria*)
2. Inconsistency (*or heterogeneity*)
3. Indirectness (*PICO and applicability*)
4. Imprecision
5. Publication bias

3 factors can increase quality

1. large magnitude of effect
2. opposing plausible residual bias or confounding
3. dose-response gradient



RevMan と GRADEpro

いずれもエビデンスの統合、質評価などに使うフリーソフトである。



Review Manager (RevMan): コクラン システムティックレビューの作成および維持のために使われるソフトウェア (version 5.1, 2012年4月時点)。

RevMan を使用してシステムティックレビュー プロトコルの執筆および管理に加え、本文、表、研究データを含むレビューの完成が可能となる。入力されたデータのメタアナリシスを実施し、結果を図で提示できる。

RevManのwebサイト (<http://ims.cochrane.org/revman>) を参照。



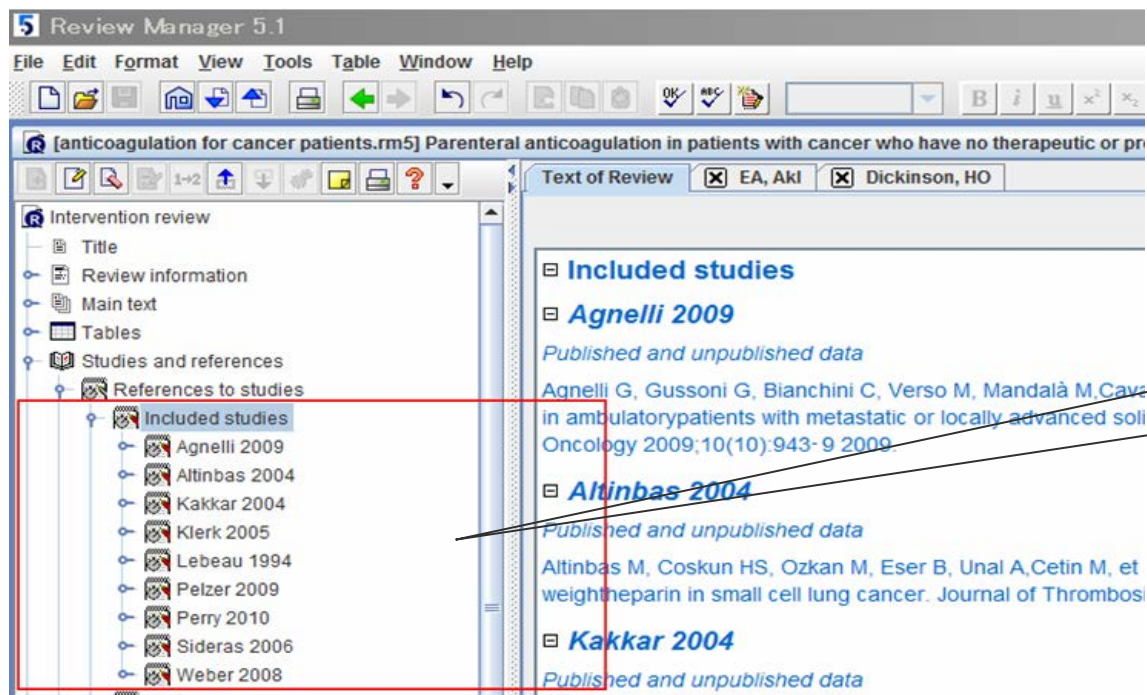
GRADE profiler (GRADE pro) : エビデンス・プロファイルやSoF テーブルなどを作成、管理、共有するためのアプリケーションで、ウェブサイトからダウンロード可能である。そのヘルプファイルがGRADE ハンドブックであり、version3.2 (3.5.1更新) の内容は相原らが和訳しウェブ上に一般公開している。

- http://homepage3.nifty.com/aihara/grade_handbook3.2.html)
- [Online GRADE ハンドブック\(日本語\)](#)

エビデンスの統合

Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation (Review)

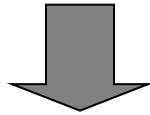
Akl EA, Gunukula S, Barba M, Yosuico VED, van Doormaal FF, Kuipers S, Middeldorp S, Dickinson HO, Bryant A, Schünemann H



Title/abstract screening、full text screeningの後、レビューでは、9件のRCTが組み込まれた

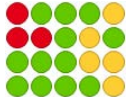
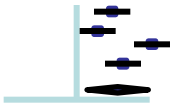
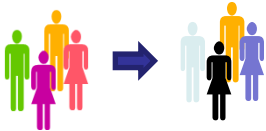
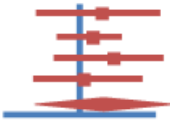
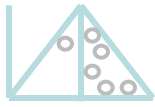
総体エビデンス (body of evidence) の質

- ・研究デザインはRCTなので、エビデンスの質を“High”としてスタートする。
- ・グレードダウンの5要因はどうか？



アウトカムごとに評価する！

- Risk of bias(*)
- Inconsistency
- Indirectness
- Imprecision
- Publication bias

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication bias)
				

(*) study limitationsと同じである

バイアスのリスク (risk of bias)

バイアス(bias)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。バイアスの程度を把握することは困難であるため、バイアスのリスク(risk of bias)として検討したことを明示する。

バイアスのリスク (risk of bias) (*)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
<p>RCT:</p> <ol style="list-style-type: none"> Lack of allocation concealment Lack of blinding Incomplete accounting of patients and outcome events Selective outcome reporting bias Other limitations (stopping early for benefit, etc) 			

これらの基準について Risk of bias (RoB)を評価する。アウトカムごとに個々の研究を評価。

	Random sequence generation	Allocation concealment	Blinding of participants	Blinding of providers	Blinding of data collection	Blinding of outcome assessment	Incomplete outcome data	Selective reporting (reporting bias)	Other bias	Intention-to-treat analysis
Agnelli 2009	●	●	●	●	●	●	●	●	●	●
Allinbas 2004	●	●	●	●	●	●	●	●	●	●
Kakkar 2004	●	●	●	●	●	●	●	●	●	●
Klerk 2005	●	●	●	●	●	●	●	●	●	●
Lebeau 1994	●	●	●	●	●	●	●	●	●	●
Pelzer 2009	●	●	●	●	●	●	●	●	●	●
Perry 2010	●	●	●	●	●	●	●	●	●	●
Sideras 2006	●	●	●	●	●	●	●	●	●	?
Weber 2008	●	●	●	●	●	●	●	●	●	●

アウトカムごとに、採用研究についてRoBを評価する

バイアスのリスク (risk of bias)

GRADE におけるエビデンスの質の評価ポイント

1. エビデンスの質は、アウトカムごとに判断する。
2. すべてのアウトカムについて、もとなる個別研究ごとに研究の限界 (limitations, risk of bias) を評価する。
3. エビデンスの質を評価する際には、各研究の質を単に平均するのではなく、個別の研究の質と結果を慎重に考慮する。
4. 研究の限界が最終的な効果推定値に寄与する程度を丁寧に考慮すべきであり、通常は、サンプルサイズやアウトカムのイベント発症数をチェックする(例、RCT でイベント数50以下の研究は除外)。
5. ほとんどの採用研究がわずかな限界しかないならば、システマティック・レビューの作成者は深刻な限界のある(含めることがエビデンスの質を大きく低下させる)研究は除外すべきである。

「診療ガイドラインのためのGRADEシステム」 p46、あるいは

<http://homepage3.nifty.com/aihara/GRADEproHelp/Study%20limitations.htm>

RCTのlimitations主要判定基準 (GRADE)

限界(Limitations)	説明
割り付けの隠蔽(Allocation concealment)がない	<ul style="list-style-type: none"> 患者を組み入れる担当者が、次に組み入れる患者がどの群に属するかを知っている場合に起こる。割り付けを、曜日、誕生日、カルテ番号などで行うときに起こりやすい。擬似ランダム化比較試験や準ランダム化比較試験で特に大きな問題となる。
盲検化(Blinding)されていない	<ul style="list-style-type: none"> 患者、介護者、アウトカムの報告者や評価者、またはデータ解析担当者が、患者がどの群に割り付けられているのを知っている。
患者やアウトカムイベントの不完全な検討(Incomplete accounting)	<ul style="list-style-type: none"> ITT解析の原則を掲げながらも、追跡からの脱落者に対してその原則を遵守していない。
選択的アウトカム報告バイアス(Selective outcome reporting bias)	<ul style="list-style-type: none"> 研究計画書に記載されているにもかかわらず、報告しているアウトカムと報告していないアウトカムとがある。
その他の限界(Other limitations)	<ul style="list-style-type: none"> 利益があったとして試験を早期中止する(stopping early for benefits)。 “患者にとって重要なアウトカム”が、妥当ではない。 クロスオーバー試験における持ち越し(carry-over)効果がある。 クラスターランダム化比較試験における組み入れバイアスがある。

観察研究のlimitations主要判定基準 (GRADE)

限界(Limitations)	説明
適切な適格基準(Eligibility criteria)を確立していない、あるいは適用していない(対照群の組み入れ)	症例対照研究の対照群選定の際に、過小(アンダー)または過大(オーバー)マッチングになっている。 コホート研究において、曝露した人と曝露していない人が背景が異なる集団から選出されている。
曝露およびアウトカムの双方における測定の不備(Flawed measurement)	曝露やアウトカムの測定が不確かな場合(例:症例対照研究における思い出しバイアス)。 コホート研究で、曝露群と非曝露群で曝露内容やアウトカムの調査方法が異なっている。
交絡(Confounding) が十分にコントロールされていない	コホート研究で、すべての既知の予後因子を測定していない、もしくは精確に測定していない。 曝露群と比曝露群で予後因子や背景因子が一致していない、または解析の際にこれらの統計学的な調整がされていない。
追跡が不十分または観察期間が短すぎる(Follow-up)	

本資料における臨床シナリオでは
RCTのため、使用しない

バイアスのリスク (risk of bias)

☐ Risk of bias table

RevMan5.1

Bias	Authors' judgement
Random sequence generation (selection bias)	Low risk
Allocation concealment (selection bias)	Low risk
Blinding of participants and personnel (performance bias)	Low risk
Blinding of providers	Low risk
Blinding of data collectors	Low risk
Blinding of outcome assessment (detection bias)	Low risk
Blinding of data analysts	High risk
Incomplete outcome data (attrition bias)	Low risk
Selective reporting (reporting bias)	Low risk
Other bias	Low risk
Intention-to-treat analysis	Low risk

コクランの risk of bias基準は、基本的に6つのドメイン。

- Random sequence generation.
- Allocation concealment
- Blinding of participants and personnel
 - Blinding of outcome assessment
- Incomplete outcome data
- Selective reporting
- Other sources of bias

Cochrane-handbook (chapter 8.5: risk of bias)

バイアスのリスク (risk of bias)

アウトカム：死亡率

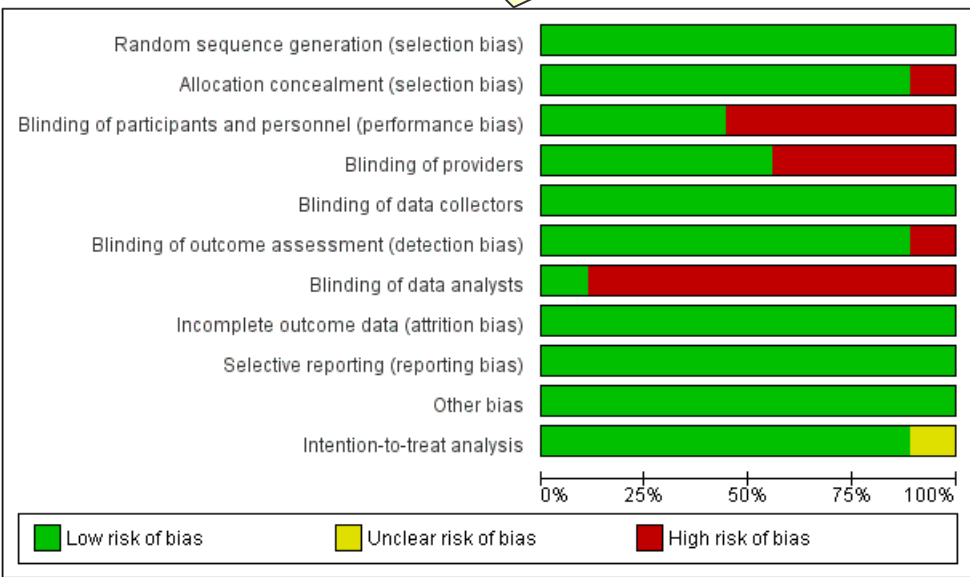
Figure 1. 'Risk of bias' summary - mortality at 12 months: review authors' judgements about each methodological quality item for each included study that evaluated mortality

個々の研究の risk of biasを評価する (within study)

	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	+	+	-	-	+	+	-	+	+	+	+

縦横の表示
を変えるだけ

アウトカムについて、複数の研究を対象としてrisk of biasを評価する。(across studies)



↓ どのように判定するのか？

バイアスのリスク (risk of bias)

Risk of bias評価は4つのレベルで実施する。

Cochrane-handbook (chapter 8.7: risk of bias)

1. Summarizing risk of bias for a study across outcomes
2. Summarizing risk of bias for an outcome within a study (across domains)
3. Summarizing risk of bias for an outcome across studies (e.g. for a meta-analysis)
4. Summarizing risk of bias for a review as a whole (across studies and outcomes)

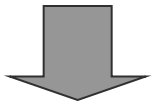
解説:

1. ある一つの研究を対象に、全てのアウトカムのrisk of biasを評価する。一つの研究内で異なるアウトカムでは、risk of biasが異なってくる場合がある。
2. ある一つの研究内の(あらゆる領域で)一つのアウトカムのrisk of biasを要約する。異なるアウトカムではrisk of biasが異なる場合があるので、一つの研究内のrisk of biasを要約する場合に推奨される。一つのアウトカムのrisk of bias要約評価には、そのアウトカムに関連する全ての項目(割付の順序の隠蔽などの研究レベルの項目、ならびに盲検化などのアウトカム固有の項目)を含めるべきである。
3. 複数の研究(メタアナリシスなど)を対象に、一つのアウトカムのrisk of biasを要約する。レビューの著者が行う主要な要約評価がこれに当たり、この評価は、第11章(11.5項)の説明にあるとおり、“SoFテーブル”の中の“エビデンスの質”の判定に盛り込まれる。以下の説明にもあるように、risk of biasの高い試験結果をメタアナリシスに含めると、そのような試験を除外した場合と比較し、エビデンスの質が低くなる。
4. ある一つのレビュー全体(複数の研究と複数のアウトカムの全て)を対象に、risk of biasを要約する。複数の研究やアウトカムのエビデンスの全体的risk of biasを判定するのは、診療ガイドラインなどの特定の状況に限定すべきであり、さまざまな場面での意思決定の支援を目的とするシステムティック・レビューの場合はこのような判定をすべきではない。

バイアスのリスク (risk of bias)

レビューの著者は、一つの研究内の重要なアウトカムのrisk of bias、ならびに複数の研究全体の重要なアウトカムのrisk of biasの双方について、明瞭な判定を下すべきである。そのためには、これらの要約評価の鍵となる(複数の)最重要領域(主要領域 "key domain")を特定する必要がある。

(あるアウトカムについて、low risk of bias, unclear risk of bias, high risk of biasの研究があるので、“判定不能”という判断をしてはならない)



研究別の6つのrisk of bias基準に基づき、ある一つの特定のアウトカムを取り上げた一連の研究全体のrisk of bias (risk of bias for an outcome **across** studies)を勘案してエビデンスの質のグレードダウン判定を行うのは容易ではないが、GRADEは”**risk of bias評価の指針**”を提案している。

バイアスのリスク (risk of bias) (GRADE)

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)

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。

バイアスのリスク (risk of bias)

RCT における「研究の限界」(risk of bias) の評価

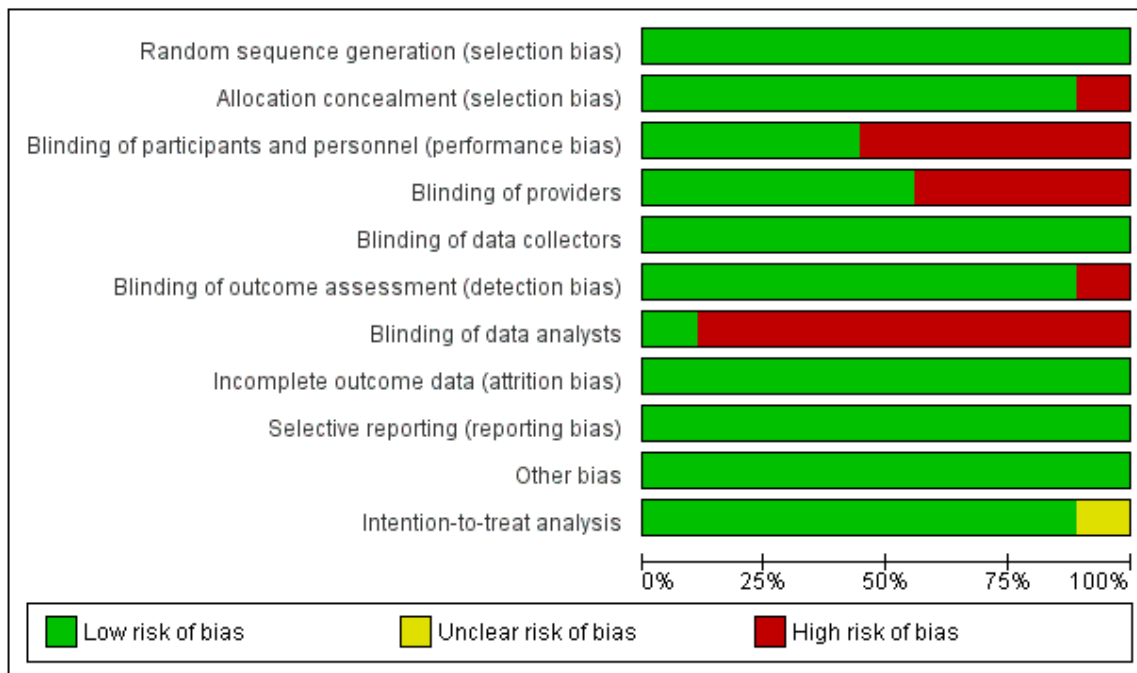
Risk of bias / 限界(limitations)	個別研究 (Within a study)	複数の研究全般 (Across studies)	複数の研究全般に関する解釈 * (Interpretation across studies)
No serious limitations 深刻な限界はなく、グレードは下げない	すべての主要基準(表1.5.1-2, 47頁)に関して、Low risk of bias である。	ほとんどの情報が「Low risk of bias」研究から得られている	エビデンスの質「高」: 真の効果が効果推定値の近くに存在する。
Serious limitations 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」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。

この評価は、GRADEに限らずシステマティック・レビューの質評価の基本である。

「診療ガイドラインのためのGRADEシステム」 p51 参照

バイアスのリスク (risk of bias)

アウトカム : 死亡率



risk of bias across studies

例) コクラン risk of bias基準の“Blinding of data analysis”項目は、赤色のhigh risk of biasが、89% (8/9件)である。

死亡率というアウトカムを考えると、“ほとんどの情報が low risk of bias研究から得られている”と評価できる。

従って、本アウトカムに関する、研究全般にわたるrisk of biasという評価においては、**エビデンスの質を下げない。**
(no serious risk of bias)

バイアスのリスク (risk of bias)

アウトカム : QOL

Figure 7. 'Risk of bias' summary - quality of life: 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?
Sideras 2006	+	+	-	-	-	-	-	+	+	+	?

アウトカムが異なると、研究の数が異なり、risk of biasも異なってくる可能性がある。

High risk of biasと判断できる。
(11項目中、6項目が high~unclearである)

総体エビデンス (body of evidence) の質

アウトカム : 全死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias				

Vast majority of studies had allocation concealment, and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.

アウトカム : QOL

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
Very serious risk of bias (-2)			serious imprecision (-1)	

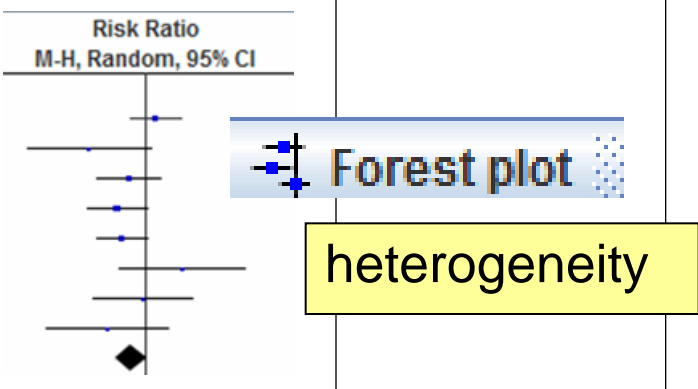
High risk of bias and only 138 patients enrolled.

これらの評価がEvidence profileに表示される。
各要因に関する判断を脚注で説明することが重要。

結果の非一貫性 (inconsistency)

非一貫性(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。異質性があるにも関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。

アウトカム： 死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias				

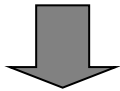
結果の非一貫性 (inconsistency)

非一貫性の4要因:

- 集団(例、重症の集団でより薬剤効果が高い)
- 介入(例、薬剤の用量が多いほど効果が大きい、共介入、比較介入が異なる)
- アウトカム(例、追跡期間が長いほど治療効果が減少する)
- 研究方法(例、RCTにおいて、risk of biasが高いものと、低いもので効果が異なっている)

もし、非一貫性の要因が“研究方法”である場合には、risk of bias が低い研究のみによる効果推定値を提供すべきである。

「結果が非一貫性」とであると評価するための、4つの基準がある。



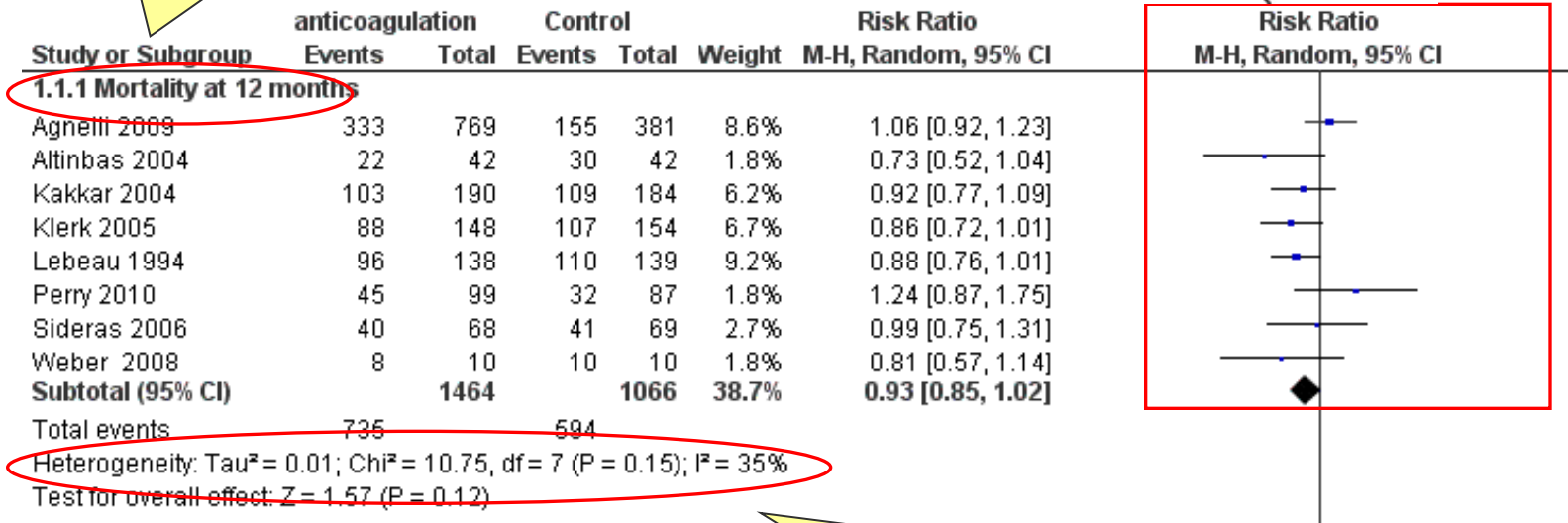
1. 点推定値が研究間で異なり、その相違がかなり大きい。
2. 各信頼区間の重なりが、ほとんどまたは全くない。
3. Cochrane Q: 異質性の検定(二択の帰無仮説: 全研究で差がない)で、p 値が小さい。
4. I^2 値(研究間の異質性を示す)が高い。

結果の非一貫性 (inconsistency)

RevManの forest plotをみる

アウトカム：死亡率(12ヶ月後)

Forest plot



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。

総体エビデンス (body of evidence) の質

アウトカム： 死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias	serious inconsistency (-1)			

There is moderate heterogeneity among studies included in the analysis of death at 12 months (I²=41%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.

これらの評価がEvidence profileに表示される。
各要因に関する判断を脚注で説明することが重要。

エビデンスの非直接性(indirectness)

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、適用性 applicabilityと同義)と同じである。

アウトカム： 死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias	serious Inconsistency (-1)	<ul style="list-style-type: none"> •Population •Intervention •Comparison •Outcome 		

4つのタイプの indirectness を検討する

エビデンスの非直接性(indirectness)

4つのタイプの非直接性

1. 研究対象集団の違い (applicability)
2. 介入の違い (applicability)
3. 比較の違い
4. アウトカム測定の違い (surrogate outcomes)

たとえば、非直接的な比較とは、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。

本アウトカムに関する、研究全般にわたるindirectness評価は、**エビデンスの質を下げない。**
(no serious indirectness)

総体エビデンス (body of evidence) の質

アウトカム： 死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias	serious inconsistency (-1)	no serious indirectness		

これらの評価がEvidence profileに表示される。
各要因に関する判断を脚注で説明することが重要。

データの不精確さ(imprecision)

サンプルサイズやイベント数が少なく、そのために効果推定値を取りまく信頼区間が幅広いとき、結果は不精確である。

アウトカム：死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias	serious Inconsistency (-1)	no serious indirectness	<ul style="list-style-type: none"> •サンプルサイズ •イベント数 •信頼区間 	

基本的には、これらの基準を使って imprecision を評価する

データの不精確さ (imprecision)

GRADEシステムにおいては、システマティック・レビューと診療ガイドラインとでエビデンスの質の定義が異なる。

システマティック・レビュー：
「効果推定値が正しいという確信の程度」

診療ガイドライン：
「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」

この両者間の定義の違いが、もっとも顕著に表れるのが、Imprecisionの評価である。

GRADEハンドブック(ver. 3.2 相原_翻訳版)の imprecisionも参照のこと。
<http://homepage3.nifty.com/aihara/GRADEproHelp/Imprecision.html>

データの不精確さ (imprecision)

通常、信頼区間はエビデンスの質におけるランダム誤差の影響を示唆する、わかりやすい指標である。信頼区間(通常95%)は、95%の確率で真実がその中のどこかに存在すると考えられる範囲である。

RevMan5.1

Study or subgroup	Heparin n/N	Control n/N	Risk Ratio M- H,Random,95% CI	Weight	Risk Ratio M- H,Random,95% CI
Agnelli 2009	333/769	155/381		19.7 %	1.06 [0.92, 1.23]
Perry 2010	45/99	32/87		6.0 %	1.24 [0.87, 1.75]
Weber 2008	8/10	10/10		6.1 %	0.81 [0.57, 1.14]
Klerk 2005	88/148	107/154		16.9 %	0.86 [0.72, 1.01]
Kakkar 2004	103/190	109/184		16.0 %	0.92 [0.77, 1.09]
Sideras 2006	40/68	41/69		8.6 %	0.99 [0.75, 1.31]
Altinbas 2004	22/42	30/42		6.1 %	0.77 [0.52, 1.04]
Lebeau 1994	96/138	110/139		20.6 %	0.88 [0.76, 1.01]
Total (95% CI)	1464	1066		100.0 %	0.93 [0.85, 1.02]

信頼区間(CI)は多くの場合、不精確さの度合いを反映しているが、
CIがどの程度狭いと確信をもてるか？

信頼区間は、サンプルサイズやイベント数によって大きく変化することに注意。

データの不精確さ (imprecision)

システマティック・レビューにおける(2値変数の)不精確さの評価

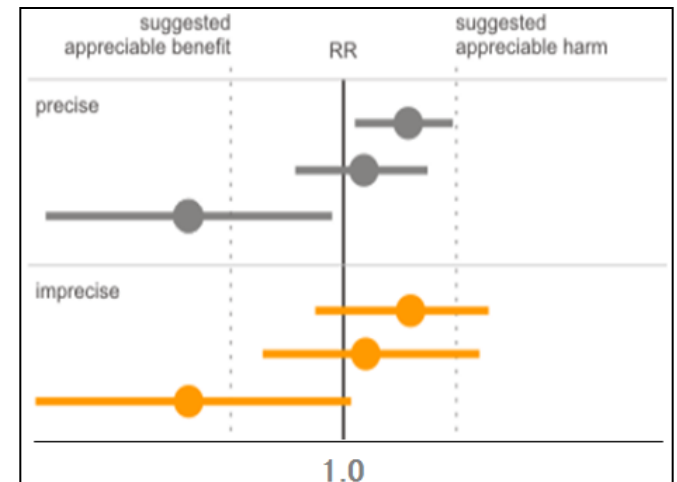
サンプルサイズが非常に大きくない限り(患者数が少なくとも2000名、おそらく4000名)、OISの基準を満たさないなら、不精確としてグレードを下げる。



もしOIS基準を満たし、かつ95%信頼区間が“効果なし”を含まない(RRの信頼区間が1を含まない)ならば、精確であると判定する



もしOIS基準を満たし、信頼区間が“効果なし”を含み、かつ、信頼区間が重要な利益または重要な害の閾値を越える場合は、不精確としてグレードを下げる。

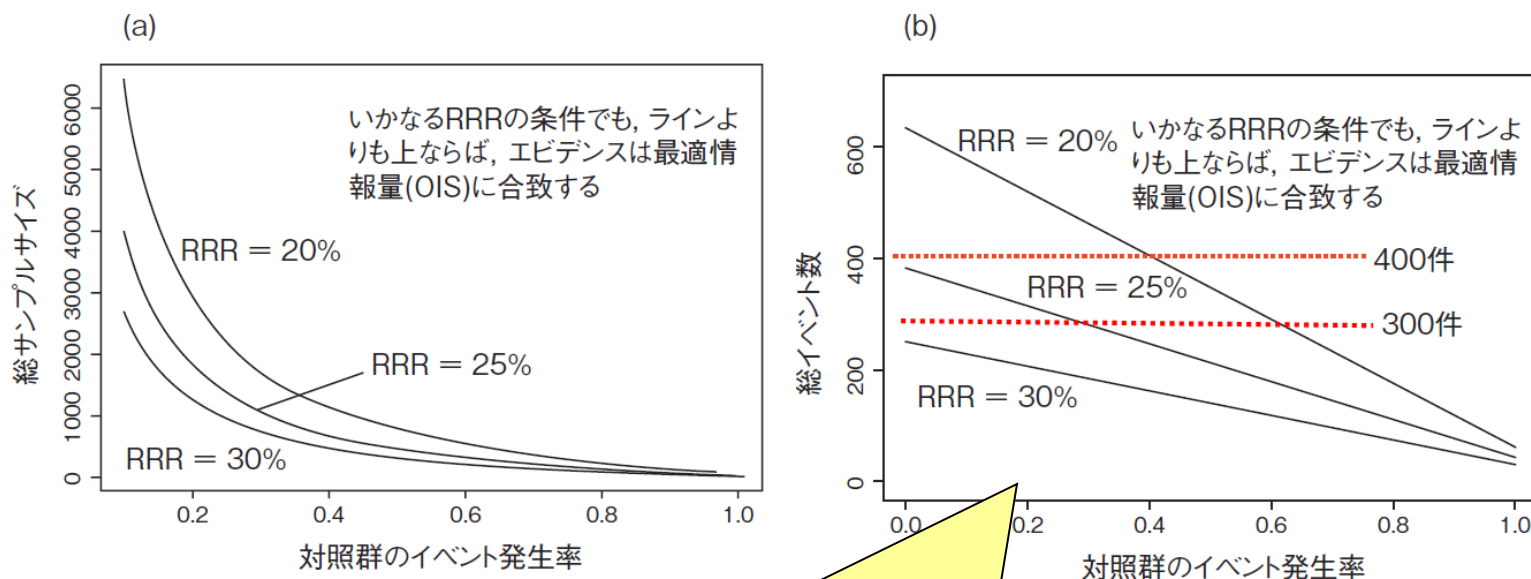


(OIS: optimal information size 最適情報量)

データの不精確さ (imprecision)

最適情報量 OIS (Optimal information size)

図 1.5.4-2 対照群イベント発生率と RRR を変えた場合の総サンプルサイズ (左図 a) および総イベント数 (右図 b)

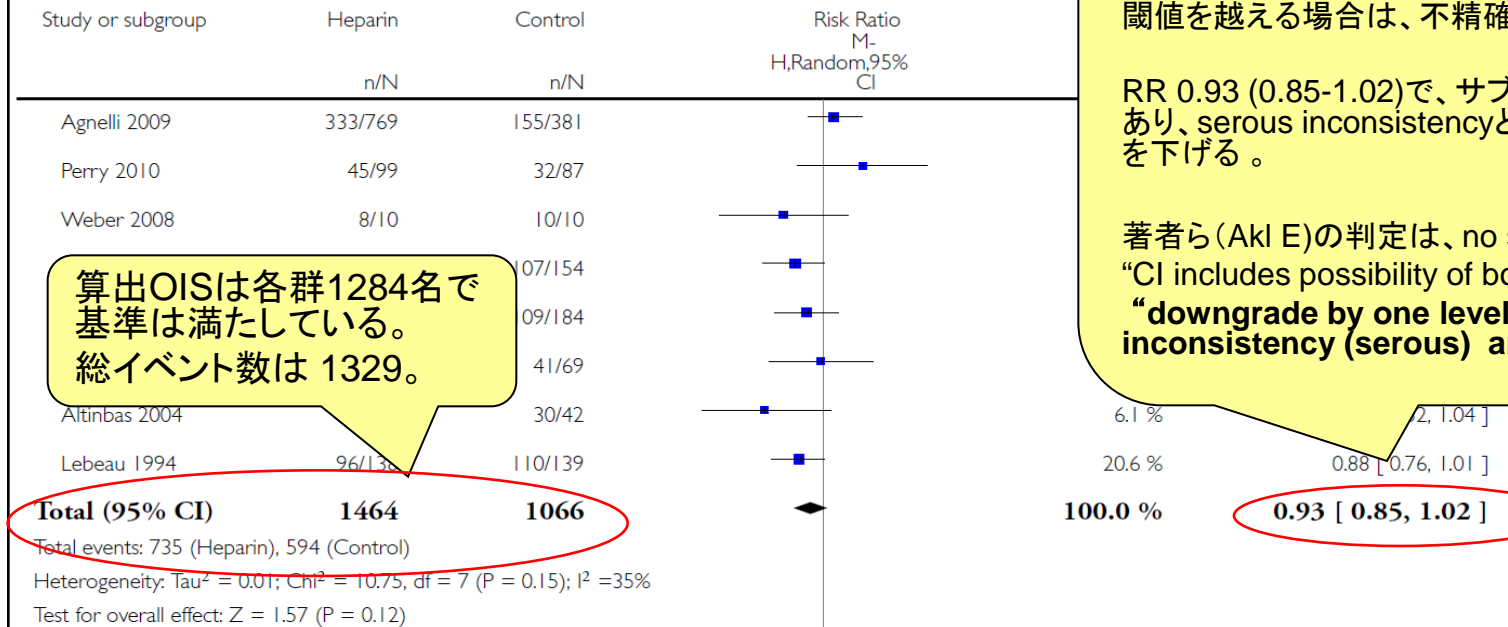


例えば、不精確さを理由に評価を下げるべき閾値としては、対照群イベント率 25% の場合、RRR 25% 以上の効果を得るためには、300 件以上のイベント数が合理的であることを示している。

データの不精確さ(imprecision)

CQ-1

アウトカム：死亡率



算出OISは各群1284名で
基準は満たしている。
総イベント数は 1329。

「もしOIS 基準を満たし、信頼区間が“効果なし”のみ、かつ、信頼区間が重要な利益または重要な害の閾値を越える場合は、不精確としてグレードを下げる。

RR 0.93 (0.85-1.02)で、サブ解析では統計的有意性あり、serious inconsistencyとあわせて1段階グレードを下げる。

著者ら(Akl E)の判定は、no serious imprecision
“CI includes possibility of both harms or benefits”
“downgrade by one level when considering inconsistency (serious) and imprecision”.

従って、レビューの本アウトカムのエビデンス (imprecision項目) の質は、非一貫性 (inconsistency) との総合的な評価として1段階下げる。

データの不精確さ (imprecision)

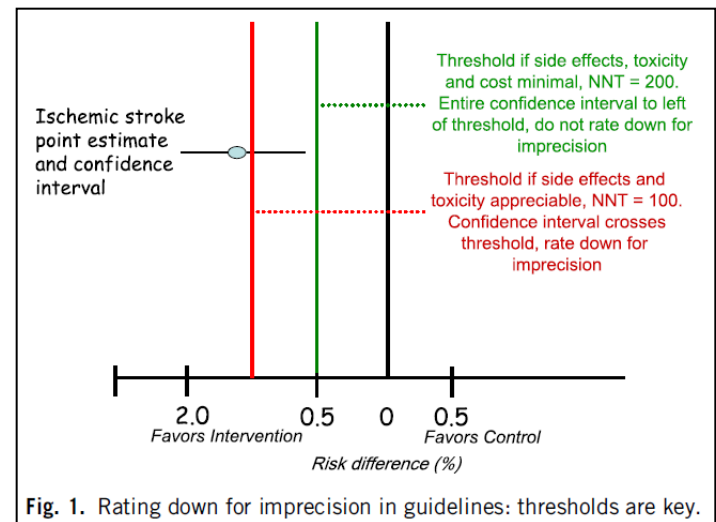
診療ガイドラインにおける(2値変数の)不精確さの評価

信頼区間が、治療を推奨するかしないかの臨床決断閾値をまたぐか？
もし、またぐならば、不精確としてグレードを下げる。



もし閾値をまたがないならば、OIS 基準を満たすか。あるいは、イベント発生率が非常に低く、かつサンプルサイズが非常に大きい(患者数が少なくとも2000名、おそらく4000名)か？
CIの閾値に関する基準もOIS基準もどちらの基準も満たさないなら、不精確としてグレードを下げる。

(OIS: optimal information size 最適情報量)



総体エビデンス (body of evidence) の質

アウトカム： 死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious risk of bias	serious inconsistency (-1)	no serious indirectness	no serious imprecision	

CI interval includes effects suggesting benefit as well as no benefit.

これらの評価がEvidence profileに表示される。
各要因に関する判断を脚注で説明することが重要。

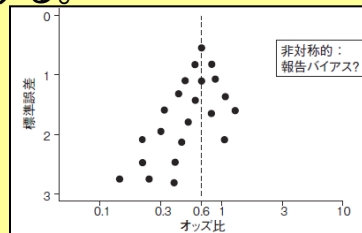
出版バイアス(publication bias)

出版バイアスは、研究が選択的に出版されることによって、根底にある有益または有害な効果が系統的に過小評価または過大評価されることをいう。

アウトカム：死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication bias)
no serious risk of bias	serious Inconsistency (-1)	no serious indirectness	no serious imprecision	<ul style="list-style-type: none"> •Funnel plot •Fail-Safe N •trim-and-fill 法 •他

基本的には、これらの基準を使って出版バイアスを評価する。



出版バイアス(publication bias)

報告バイアスには2つのカテゴリがある。

1. アウトカム報告バイアス(著者あるいは研究スポンサーが、事前に設定したアウトカムをすべて報告せずに、自身にとって都合のよいアウトカムデータのみを報告する場合をいう)で、個々の研究内で生じるものである。このため、選択的アウトカム報告バイアスの評価は、GRADEでは、個々の研究のrisk of biasの評価基準に含まれる。
2. 出版バイアス(publication bias)で研究全体が報告されていない場合をいう。

出版バイアスを疑うべき典型的な状況は、発表済みのエビデンスが少数の試験に限定され、そのいずれもが研究介入の利益を示しているときに起こる。

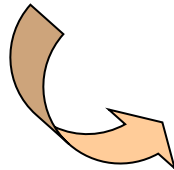


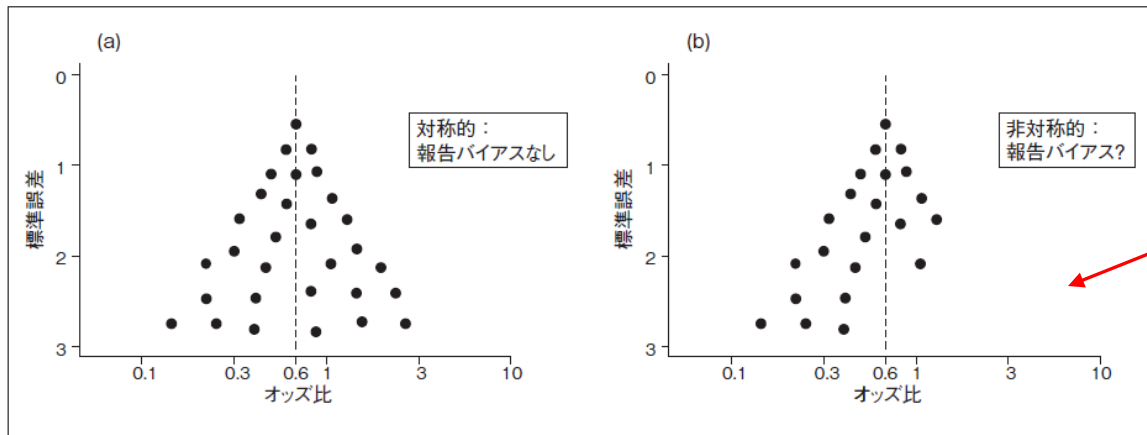
表 20.1-3

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、図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

出版バイアス(publication bias)

図 1.5.5-1 ファンネルプロット



(b) では右下側に相当する研究がプロットされず、出版バイアスがあることを示唆している。

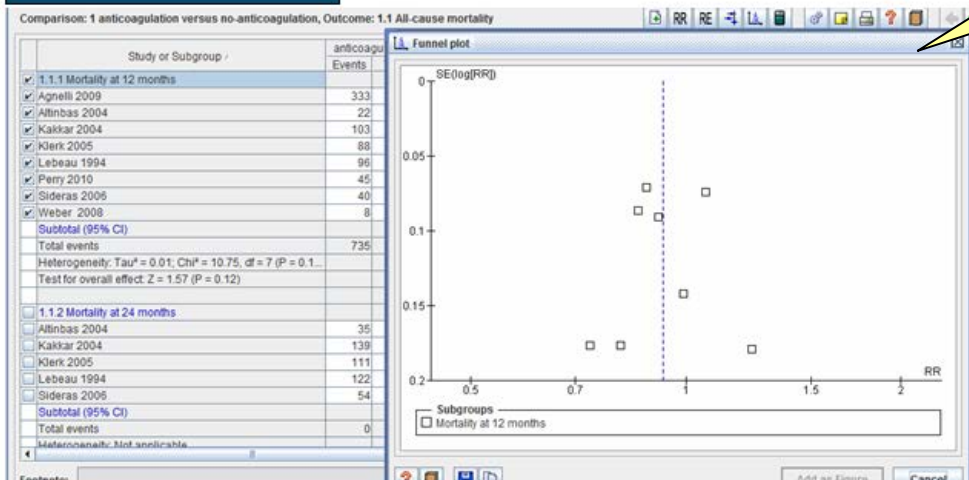
ファンネルプロットによって、左右対称性の検定をプロットの目視確認に限定することは、一般的に行われていることではあるが、あまりにも誤差の影響を受けやすい。

グレードダウン

出版バイアス(publication bias)

アウトカム：死亡率

RevMan5.1



The inverted funnel plot not suggest publication bias

ファンネルプロットの統計的解析では...

リスク比公開バイアス検出の有意水準 (初期値 0.10)	0.10
研究数	8
1. Beggの方法	
Kendall相関係数	-0.0714
Z値	-0.2474
p値	0.8046
公開バイアスなしの仮定は棄却されなかった	

本アウトカムについて、publication biasは、検出されなかった(not detected)。

グレードアップの3要因はどうか？

RCTのエビデンスの質は「高」、観察研究のエビデンスの質は「低」から評価を開始する。しかし、稀ながら、研究結果によっては、観察研究(2件以上)のエビデンスの質の評価を、「中」あるいは「高」に上げることもある。エビデンスの質の評価を上げる主たる判断要因には3つはある。

- Large magnitude
- Dose response gradient
- confounders

効果の程度が大きい (large magnitude) (*)	用量反応勾配 (dose response gradient)	交絡因子 (confounders)
<ul style="list-style-type: none"> ● 大きい(large) RR >2 または0.5 ● 非常に大きい(very large) RR >5 または 0.2 	用量反応勾配の存在は、観察研究の結果における確信を高めるため、エビデンスの質を上げることがある。	観察研究またはランダム化試験における全ての交絡因子が、「明示された効果を減少させる方向に働く」、あるいは、「効果が観察されていないのに効果を増大させる方向に働く」ことがある。
ただし、グレードを上げることができるのは、研究の妥当性に問題がない(なんらかの理由で評価が下げられていない)観察研究に限られる。		

(*) 「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こと。

Bradford Hill criteria と GRADE

Table 1 Bradford Hill criteria of causality and their relation to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria for upgrading and downgrading

Bradford Hill criteria	Consideration in GRADE
Strength	Strength of association and imprecision in effect estimate
Consistency	Consistency across studies, ie, across different situations (different researchers)
Temporality	Study design, specific study limitations; RCTs fulfil this criterion better than observational studies, properly designed and conducted observational studies
Biological gradient	Dose—response gradient
Specificity	Indirectness
Biological plausibility	Indirectness
Coherence	Indirectness
Experiment	Study design, randomisation, properly designed and conducted observational studies
Analogy	Existing association for critical outcomes will lead to not downgrading the quality, indirectness

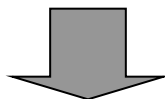
[The GRADE approach and Bradford Hill's criteria for causation](#)

[JECH 2011;65:392-395 doi:10.1136/jech.2010.119933](#) ()*

総体エビデンス (body of evidence) の質

アウトカム：死亡率

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
no serious	serious inconsistency (-1)	no serious indirectness	no serious imprecision (*)	undetected



(注) グレードアップの要因は該当するものがなかった。

最終的なエビデンスの質は、

[High] ⊕⊕⊕⊕ から1段階グレードダウンして、[Moderate] ⊕⊕⊕○



QOLについて、同じように body of evidenceの質を評価する。

総体エビデンス (body of evidence) の質

アウトカム : QOL

バイアスのリスク (risk of bias)	非一貫性 (inconsistency)	非直接性 (indirectness)	不精確さ (imprecision)	出版バイアス (publication_bias)
serious risk of bias (-2)	no serious inconsistency	no serious indirectness	serious imprecision (*)	undetected

High risk of bias and only 138 patients enrolled

(注) グレードアップの要因は該当するものがなかった。

最終的なエビデンスの質は、

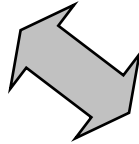
[High] ⊕⊕⊕⊕ から2段階グレードダウンして、[Low] ⊕⊕○○

大出血、小出血、症候性VTE について、同じように body of evidenceの質を評価する。

総体エビデンス (body of evidence) の質

アウトカムレベルのエビデンス (body of evidence) の要約表を作成する。

SoF (Summary of finding) テーブル、GRADE evidence profile



GRADE evidence profile

http://homepage3.nifty.com/aihara/EP_heparin_aihara_20120426.html

Author(s): aihara// Elie A Akl, Sameer K Gunukula, Frederiek F van Doormaal, Maddalena Barba, Saskia Kuipers, Saskia Middeldorp, Victor E D Yosiuco, 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 Akl EA, S Gunukula, M Barba, VED Yosiuco, FF van Doormaal, S Kuipers, S Middeldorp, HO Dickinson, A Bryant, H Schunemann. Cochrane Database Syst Rev. 2011 Apr 13;(4):CD006652].

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Heparin	No heparin	Relative (95% CI)	Absolute		
All-cause mortality (follow-up 12 months)												
8	randomised trials	no serious risk of bias ¹	serious ²	no serious indirectness	no serious imprecision ³	none	735/1464 (50.2%)	594/1066 (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)		
Symptomatic VTE (follow-up 12 months)												
7	randomised trials	no serious risk of bias ¹	no serious inconsistency	no serious indirectness	no serious imprecision	none	38/1338 (2.8%)	57/926 (6.2%)	RR 0.55 (0.37 to 0.82)	28 fewer per 1000 (from 11 fewer to 39 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
								2.9%		13 fewer per 1000 (from 5 fewer to 18 fewer)		
Major bleeding (follow-up 12 months)												
9	randomised trials	no serious risk of bias ¹	no serious inconsistency	no serious indirectness	serious ⁴	none	30/1624 (1.8%)	23/1219 (1.9%)	RR 1.3 (0.59 to 2.88)	6 more per 1000 (from 8 fewer to 35 more)	⊕⊕⊕○ MODERATE	CRITICAL
								0.7%		2 more per 1000 (from 3 fewer to 13 more)		

アウトカム

質評価

(Continued)

GRADE evidence profile

(Continued)

Minor bleeding (follow-up 12 weeks)												
7	randomised trials	no serious risk of bias ¹	no serious inconsistency	no serious indirectness	serious ⁴	none	85/1365 (6.2%)	50/980 (5.1%)	RR 1.05 (0.75 to 1.46)	3 more per 1000 (from 13 fewer to 23 more)	⊕⊕⊕○ MODERATE	IMPORTANT
								2.7%		1 more per 1000 (from 7 fewer to 12 more)		
Health related Quality of life (follow-up 12 months; measured with: the Uniscale and the Symptom Distress Scale (SDS); Better indicated by lower values; Better indicated by lower values)												
1	randomised trials	serious ^{5,6}	no serious inconsistency	no serious indirectness	serious ⁶	none	0	-	-	not pooled	⊕○○○ LOW	IMPORTANT

Footnote

- ¹ Vast majority of studies had allocation concealment, and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.
- ² There is moderate heterogeneity among studies included in the analysis of death at 12 months (I²=41%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.
- ³ CI interval includes effects suggesting benefit as well as no benefit.
- ⁴ CI includes possibility of both harms or benefits.
- ⁵ The scores for the 2 scales were similar for the 2 study groups, both at baseline and at follow-up
- ⁶ High risk of bias and only 138 patients enrolled.

1 Vast majority of studies had allocation concealment, and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.

2 There is moderate heterogeneity among studies included in the analysis of death at 12 months (I²=41%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.

3 CI interval includes effects suggesting benefit as well as no benefit.

4 CI includes possibility of both harms or benefits.

5 The scores for the 2 scales were similar for the 2 study groups, both at baseline and at follow-up

6 High risk of bias and only 138 patients enrolled.

Evidence profile (ACCP)

http://homepage3.nifty.com/aihara/EP_heparin_accp_aihara_20120426.html

Question: Should heparin vs no heparin be used for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation? Bibliography: Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation Akl EA, S Gunukula, M Barba, VED Yosuco, FF van Doormaal, S Kuipers, S Middeldorp, HO Dickinson, A Bryant, H Schunemann. Cochrane Database Syst Rev. 2011 Apr 13;(4):CD006652].											
Quality assessment							Summary of Findings				
Participants (studies) Follow up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With No heparin	With Heparin		Risk with No heparin	Risk difference with Heparin (95% CI)
All-cause mortality (CRITICAL OUTCOME)											
2530 (8 studies) 12 months	no serious risk of bias ¹	serious ²	no serious indirectness	no serious imprecision ³	undetected	⊕⊕⊕⊕ MODERATE ^{1,2,3} due to inconsistency	594/1066 (55.7%)	735/1464 (50.2%)	RR 0.93 (0.85 to 1.02)	Study population	
										557 per 1000	39 fewer per 1000 (from 84 fewer to 11 more)
										Moderate	
										645 per 1000	45 fewer per 1000 (from 97 fewer to 13 more)
High											
791 per 1000	55 fewer per 1000 (from 119 fewer to 16 more)										
Symptomatic VTE (CRITICAL OUTCOME)											
2264 (7 studies) 12 months	no serious risk of bias ¹	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ HIGH ¹	57/926 (6.2%)	38/1338 (2.8%)	RR 0.55 (0.37 to 0.82)	Study population	
										62 per 1000	28 fewer per 1000 (from 11 fewer to 39 fewer)
										Moderate	
29 per 1000	13 fewer per 1000 (from 5 fewer to 18 fewer)										

エビデンスプロファイル表示は、Quality assessment と Summary of findings から構成されているが、その様式はさまざま、治療介入と診断介入においても異なる。

本プロファイルは、リスク差をわかりやすく記載する ACCP様式である。

Summary of finding (SoF)テーブル

SoFに関するFAQの翻訳 (*)

http://www.grade-jpn.com/grade_pro-faq.html

heparin compared to no heparin for patients with cancer who have no other therapeutic or prophylactic indication for

Patient or population: patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation

Settings: outpatient

Intervention: heparin

Comparison: no heparin

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				
Mortality Follow-up: 12 months	Medium risk population		RR 0.93 (0.85 to 1.02)	2531 (8 studies)	⊕⊕⊕○ moderate ^{1,2,3}	
	649 per 1000	604 per 1000 (552 to 662)				
Symptomatic VTE Follow-up: 12 months	Medium risk population		RR 0.55 (0.37 to 0.82)	2264 (7 studies)	⊕⊕⊕⊕ high ¹	
	29 per 1000	16 per 1000 (11 to 24)				
Major bleeding Follow-up: 12 months	Medium risk population		RR 1.3 (0.59 to 2.88)	2843 (9 studies)	⊕⊕⊕○ moderate ^{1,4}	
	7 per 1000	9 per 1000 (4 to 20)				
Minor bleeding Follow-up: 12 weeks	Medium risk population		RR 1.05 (0.75 to 1.46)	2345 (7 studies)	⊕⊕⊕○ moderate ^{1,4}	
	27 per 1000	28 per 1000				

(Continued)

http://homepage3.nifty.com/aihara/sof_heparin_aihara_20120426.html

Summary of finding (SoF)テーブル

(Continued)

Health related quality of life the Uniscale and the Symptom Distress Scale (SDS); Better indicated by lower values Follow-up: 12 months	See comment	See comment	Not estimable ⁵	0 (1 study)	⊕⊕○○ low ⁶
---	-------------	-------------	----------------------------	----------------	--------------------------

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; RR: Risk ratio;

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

¹ Vast majority of studies had allocation concealment, and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.

² There is moderate heterogeneity among studies included in the analysis of death at 12 months (I²=41%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.

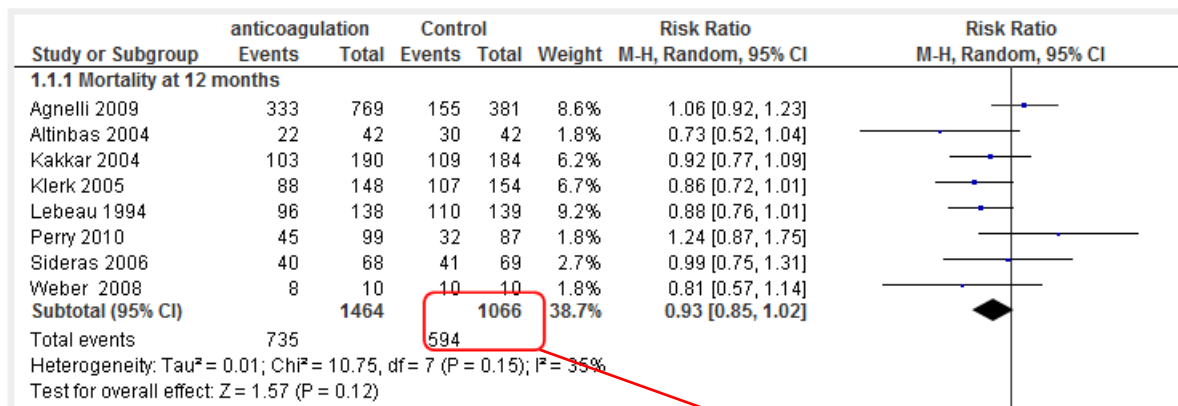
³ CI interval includes effects suggesting benefit as well as no benefit.

⁴ CI includes possibility of both harms or benefits.

⁵ The scores for the 2 scales were similar for the 2 study groups, both at baseline and at follow-up

⁶ High risk of bias and only 138 patients enrolled.

対照群リスクからの対応リスクの計算



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)

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

Outcome: All-cause mortality dichotomous continuous

pooled not pooled range of effects single study not measured not reported [?]

Number of pwith event Intervention: 735 total: 1464 50.2% [?]

with event Control: 594 total: 1066 55.7% [?]

Control risk: Low 0.0% Moderate 64.5% High 79.1% [?]

Label: Low Moderate High

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

Auto absolute effect calculation Absolute: 39 fewer per 1000 95% CI from 84 to 11 [?]

Delete Revert Go to quality assessment

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

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)
	Assumed risk	Corresponding risk			
	No heparin	Heparin			
All-cause mortality Follow-up: 12 months	Study population		RR 0.93 (0.85 to 1.02)	2531 (8 studies)	⊕⊕⊕⊕ moderate ^{1,2,3}
	557 per 1000	518 per 1000 (473 to 568)			
	Moderate				
	645 per 1000	600 per 1000 (548 to 658)			
	High				
	791 per 1000	736 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		Effect		Quality	Importance
Heparin	No heparin	Relative (95% CI)	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)		

エビデンスの質 (SR)

結論:

ヘパリンは、(24ヶ月時の死亡率の有意な減少と関連していたが) 12ヶ月時の死亡率とは関連していなかった。

ヘパリンは静脈血栓塞栓症の減少とも関連していたが、本レビューのRCTに基づくと、大出血や小出血、もしくはQOLには有意な効果はなかった。

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) ⊕⊕○○

推奨を作成する場合は、アウトカム全般に関する全体的なエビデンスの質を1つだけ決定する。
(overall quality **across** outcomes)

Part 2: エビデンスから推奨へ

システマティックレビューのための PRISMA声明



<http://www.prisma-statement.org/>

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) : 2009年6月に発表された声明で、RCTに関するメタアナリシス報告の質を向上させるためのツールである。27項目のチェックリストと、4段階のフロー図からできている。

診療ガイドラインのためのGRADEシステム、付録H: QUOROMからPRISMAへ
<http://www.grade-jpn.com/PRISMA-QUOROM-1.pdf>

Part 2

エビデンスから推奨へ

Going from evidence to
recommendations

診療ガイドライン

診療ガイドラインは通常、「医師および患者が、ある特定の状況下で適切な医療に関する意思決定を行えるように支援するための、系統的に作成された文書」として定義される。

診療ガイドラインはEvidence-Based Medicine (EBM) と呼ばれる、1991年にカナダのMcMaster大学のGordon Guyatt 医師が初めて提唱した根拠に基づく医療(EBM)を推進するためのツールである。

「診療ガイドラインのためのGRADEシステム」 p4

システマティック・レビュー

疑問の定式化
アウトカムの選択
重要度の評価

複数の研究ア
ウトカムの統合

エビデンス・プロ
ファイル作成
(GRADEprofiler)

各アウトカムに
関するエビデン
スの質を評価

RCTは「高」から、観察
研究は「低」から開始

P
I
C
O

アウトカム 重大
アウトカム 重大
アウトカム 重要
アウトカム

重要でない



Quality assessment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Study limitations																				
2. Imprecision																				
3. Indirectness																				
4. Inconsistency																				
5. Publication bias																				

高
中
低
非常に低

グレードを
下げる

1. 研究の限界
2. 非一貫性
3. 非直接性
4. 不精確さ
5. 出版バイアス

グレードを
上げる

1. 大きな効果
2. 用量反応
3. 交絡因子

各アウトカムに関する
効果推定値と結果要約
(Summary of Findings)

ガイドライン作成

推奨の作成:

- 推奨または 推奨しない(方向)
- 強いまたは 弱い(推奨度)

以下を考慮して判断:

- エビデンスの質
- 利益と害のバランス
- 価値観と好み



必要に応じて:

- 資源の利用(コスト)
- を考慮し改訂する

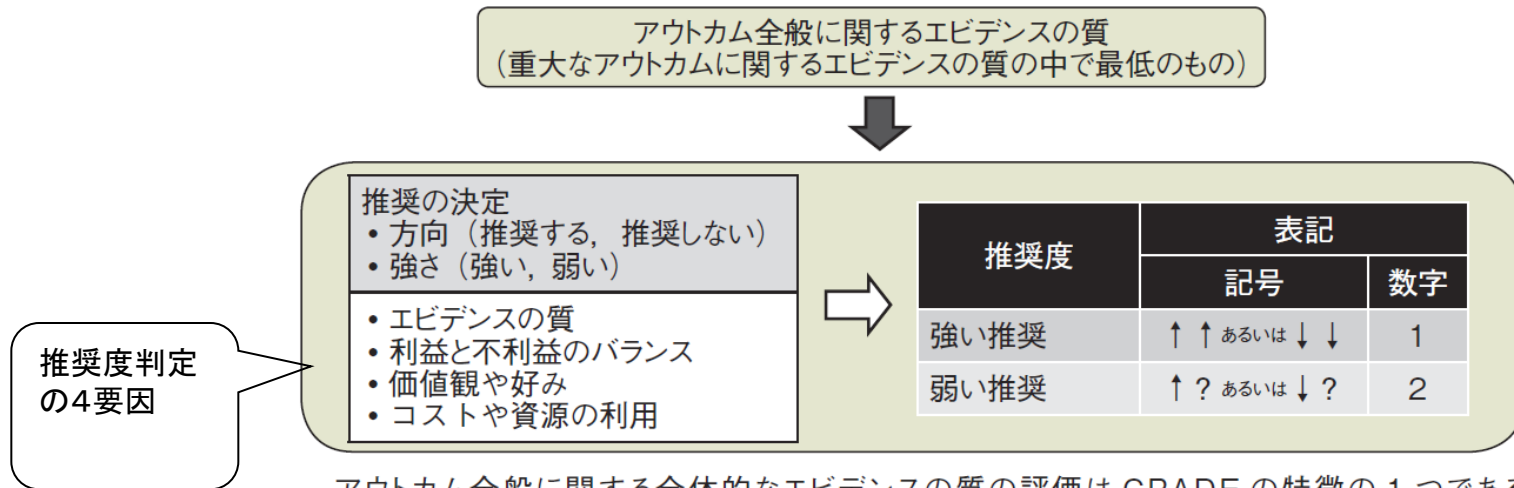


全体的なエビデンスの質
アウトカム全般に関する評価は、
重大なアウトカムに関する
エビデンスの質の中で最低のもの

- "...を使用する(強い推奨)"
- "...を使用する(弱い推奨)"
- "...を使用しない(強い推奨)"
- "...を使用しない(弱い推奨)"

推奨の強さ(推奨度)

図 3-1 GRADE による推奨度の評価



アウトカム全般に関する全体的なエビデンスの質の評価は GRADE の特徴の 1 つである (第 3 章. 3.2-1, 101 頁)。また, 推奨度は「強い」, 「弱い」 の 2 値である。

推奨の強さ(推奨度)

GRADEは、推奨(recommendation)を、「介入による望ましい効果が望ましくない効果を上回るか下回るかについて、どの程度確信できるかを示すもの」と定義する。

望ましい効果とは、その疾患の死亡率や罹患率の低下、生活の質(Quality of Life: QOL)の改善、治療の負担の軽減(治療を受ける必要性、通院の不便さの改善、など)を指し、望ましくない結果とは死亡率や罹患率の上昇、あるいはQOLへの悪影響、有害な症状の発現や検査異常の増加などを指す。

推奨の強さの分類は、「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。

強い推奨度の定義と意味

Strong recommendation

定義	介入による望ましい効果(利益)が望ましくない効果(害・負担・コスト)を上回る、または下回る確信が強い。
患者にとって	その状況下にあるほぼ全員が、推奨される行動を希望し、希望しない人がごくわずかである。
臨床医にとって	ほぼ全員(most individuals)が推奨される介入を受けべきである。ガイドラインに準じた推奨を遵守しているかどうかは、医療の質の基準やパフォーマンス指標としても利用できる。個人の価値観や好みに一致した意思決定を支援するためのフォーマルな意思決定支援は不要だろう。
政策作成者にとって	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することが可能である。

弱い推奨度の定義と意味

Weak recommendation

定義	介入による望ましい効果(利益)が望ましくない効果(害・負担・コスト)を上回る、または下回る確信が弱い
患者にとって	その状況下にある人の多くが提案される行動を希望するが、希望しない人も多い。
臨床医にとって	推奨される行動を提案し、患者が意思決定できるよう支援することは、医療の質の基準やパフォーマンス指標として利用できるだろう。意思決定支援は、患者が自身の価値観や好みに基づいて意思決定を行うのにおそらく有用だろう。エビデンスやエビデンスの要約を各自で吟味し、患者の意思決定に影響するような要因を話し合うために準備する。
政策作成者にとって	政策決定のためには、多数の利害関係者を巻き込んで実質的な論議を重ねる必要がある。パフォーマンス指標においては、管理の選択についての十分な検討が必要である。

推奨の強さ(推奨度)の表現

推奨度	表記	
	記号	数字
強い推奨	↑ ↑ あるいは ↓ ↓	1
弱い推奨	↑ ? あるいは ↓ ?	2

強い推奨の場合、GRADEワーキンググループは、「～を推奨する(recommend)」または「臨床医は～すべきである(should)」のような表現の採用を提案している。

弱い推奨の場合、GRADEワーキンググループは、「～を提案する(suggest)」または「臨床医は～かもしれない(might)」という表現の使用を提案する。

推奨の強さ(推奨度)

推奨度判定の4要因

- ① 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
- ② 望ましい効果と望ましくない効果のバランス
- ③ 価値観や好み
- ④ コストや資源の利用

GRADEはエビデンスの質と、推奨を分離している。
推奨は常に、推奨度とエビデンスの組み合わせで記述する。
(例、2B: 弱い推奨/「中」の質のエビデンス)

推奨度の要因 -① エビデンスの質

■ 診療ガイドラインにおけるエビデンスの質

The quality of evidence reflects the extent to which our confidence in an estimate of the treatment effect is adequate to support particular recommendation

「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」

診療ガイドラインでは、“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を1つだけ決定する (overall quality **across** outcomes)。

GRADEシステムには、(特徴的な) 全体的なエビデンスの質の判定基準がある。

推奨の強さの要因 -① エビデンスの質

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(*)

1. 重大(critical) なアウトカムのみを検討する。

2. 重大なアウトカム間でエビデンスの質が異なり、かつ、

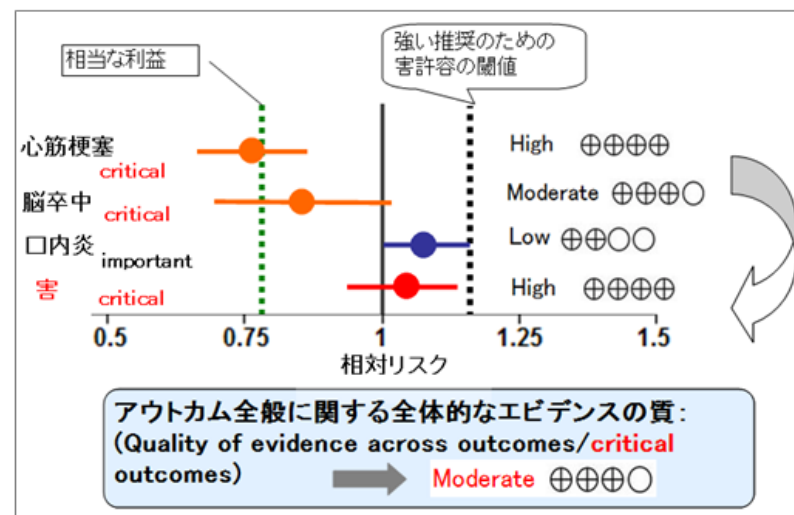
●各アウトカムが異なる方向(利益と害)を示している場合:

「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。

●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。

例:

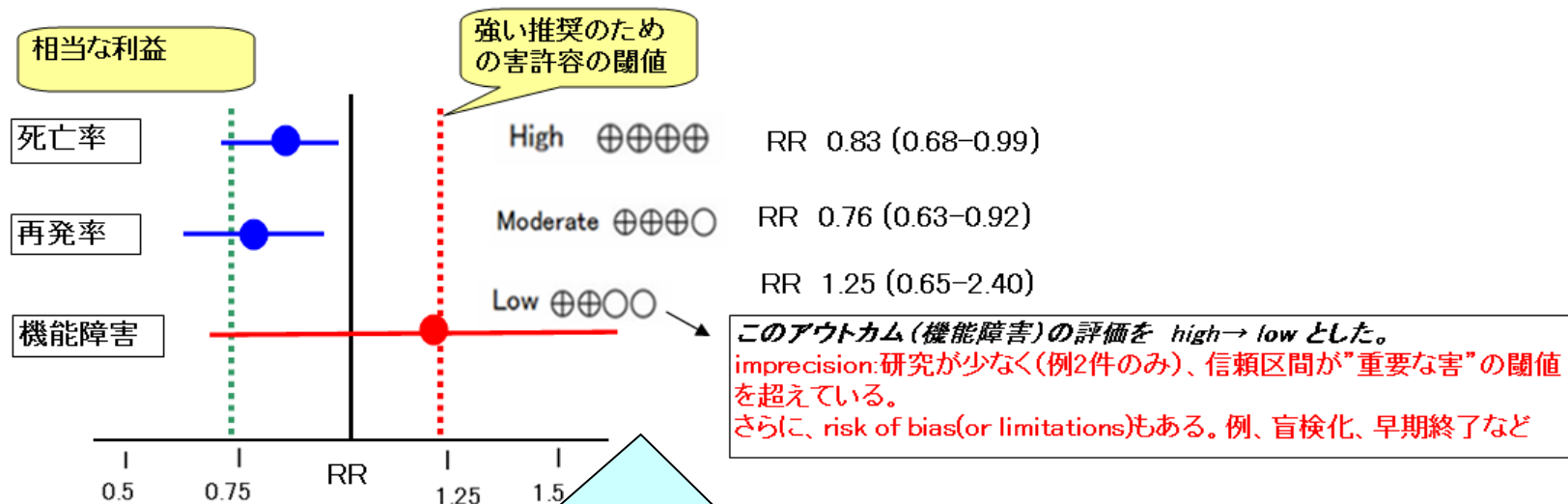


GRADEハンドブック: 全体的なエビデンスの質

<http://homepage3.nifty.com/aihara/GRADEproHelp/Overall%20quality%20of%20evidence.htm>

CPG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)



全体的なエビデンスの質としての判定は、全てのアウトカムに関して、効果が異なる方向で、特に”害”については(推奨の閾値)を超えているので、「最も低い質のエビデンスを採用」というGRADE規則により、**Low**

↓ もし、機能障害のエビデンスをみて、重要度が変更になるとどうなるか？

CPGにおけるエビデンスの質評価例-2 (仮想) : アウトカムの重要度を見直す場合:

機能障害に関する効果推定値をRR1.25ではなく、ARR 0.004、NNH 250で評価した。また、死亡率 (RR 0.83) に関するNNT 29と比較する。

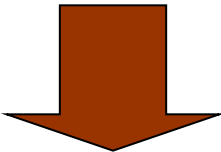
機能障害

	事象(+)	事象(-)	計
治療・曝露 (+)	a 20	b 980	1000
治療・曝露 (-)	c 16	d 984	1000
計	36	1964	2000

$\chi^2 = 0.453$ $P = 0.501$

EER = a/(a+b) = 0.02
CER = c/(c+d) = 0.02

	95%信頼区間
相対リスク RR = a/(a+b)/c/(c+d)	1.25 0.65 2.40
相対リスク減少 RRR = 1-RR	-0.25 -1.40 0.35
リスク差 RD = CER-EER	-0.004 -0.02 0.01
オッズ比 OR = (a/b)/(c/d)	1.26 0.65 2.44
治療必要数 NNT = 1/RD	-250 131 -64



死亡率

	事象(+)	事象(-)	計
治療・曝露 (+)	a 165	b 835	1000
治療・曝露 (-)	c 200	d 800	1000
計	365	1635	2000

$\chi^2 = 4.1054$ $P = 0.043$

EER = a/(a+b) = 0.17
CER = c/(c+d) = 0.20

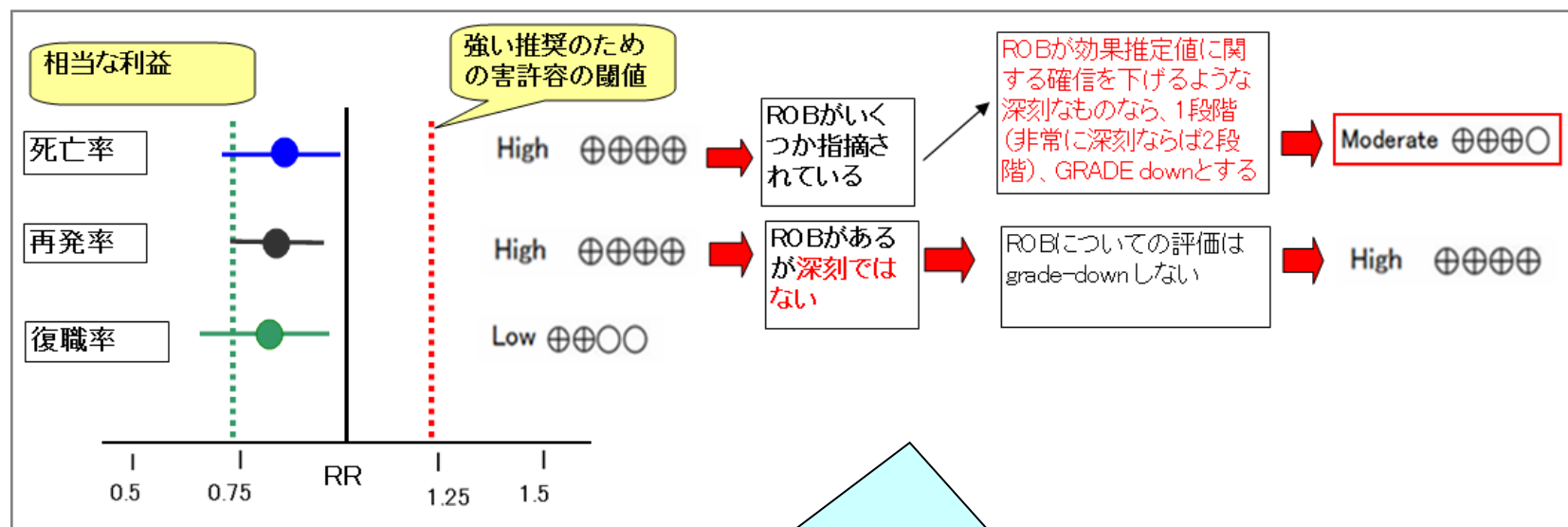
	95%信頼区間
相対リスク RR = a/(a+b)/c/(c+d)	0.83 0.68 0.99
相対リスク減少 RRR = 1-RR	0.18 0.01 0.32
リスク差 RD = CER-EER	0.04 0.00 0.07
オッズ比 OR = (a/b)/(c/d)	0.79 0.63 0.99
治療必要数 NNT = 1/RD	29 15 849

エビデンスの検索後に、アウトカム(機能障害)の重要度を1段階下げて(critical→ important)、アウトカム全般にわたるエビデンスの質評価対象から除外するかもしれない。ガイドランパネルは、以下のように最終的な推奨を記載する可能性がある。

強い推奨 / 高い質のエビデンス
 価値観や好み: この推奨は、当該介入における死亡率や再発率に高い価値を、臓器障害には低い価値を置いている。

CPG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)



全体的なエビデンスの質としての判定は、

全てのアウトカムに関して、効果が同じ方向なので、「最も高い質のエビデンスを採用」というGRADE規則により、**High** となる。

臨床疑問 (CQ)-1

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

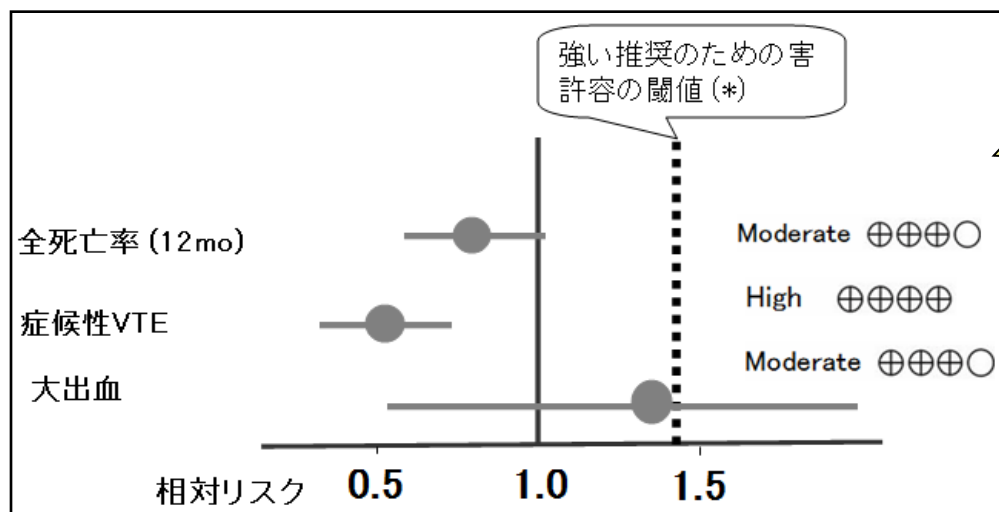
重大なアウトカムは？

- Patient**: 癌患者(肺癌、大腸癌、前立腺癌など)
- Intervention**: parenteral anticoagulation (heparin)
- Comparison**: プラセボ、no heparin
- Outcomes**:
全死亡率、大出血、症候性VTE、QOL、血小板減少

注: これらのアウトカムの重要度は、入手したエビデンスの結果により変更されることもある(*)。

推奨の強さの要因 -① エビデンスの質

GRADE evidence profileより



エビデンスの質が異なり、アウトカムが利益と害の異なる方向を示しているので、

「最も質が低いもの」を全体的なエビデンスの質とする。

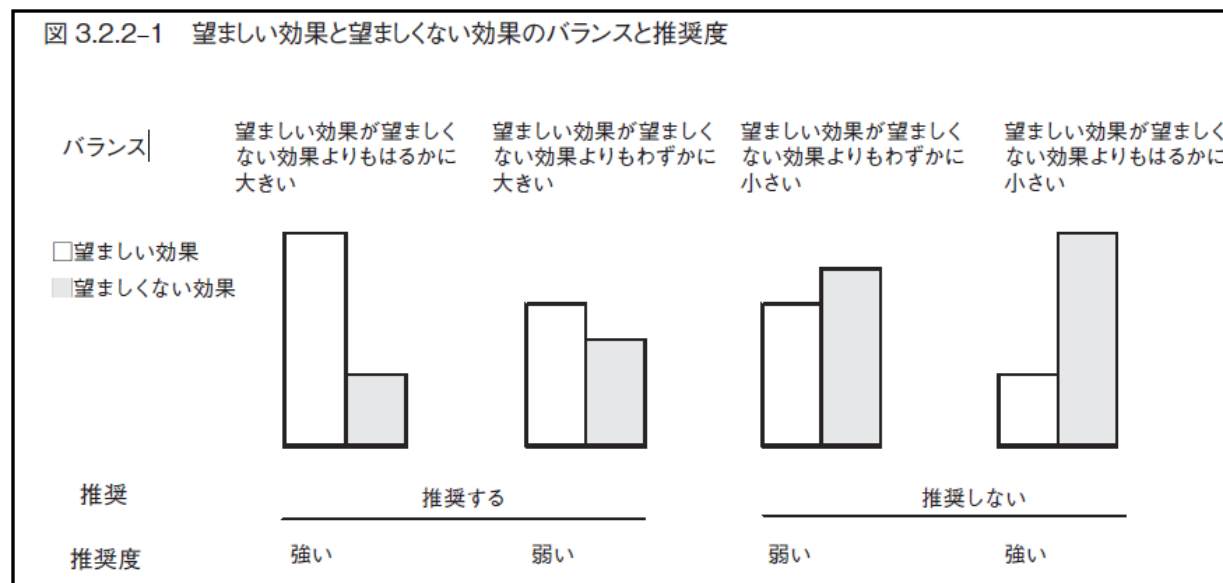
アウトカム全般に関する全体的なエビデンスの質:
(Quality of evidence across outcomes)

Moderate ⊕⊕⊕○

推奨の強さの要因 -②

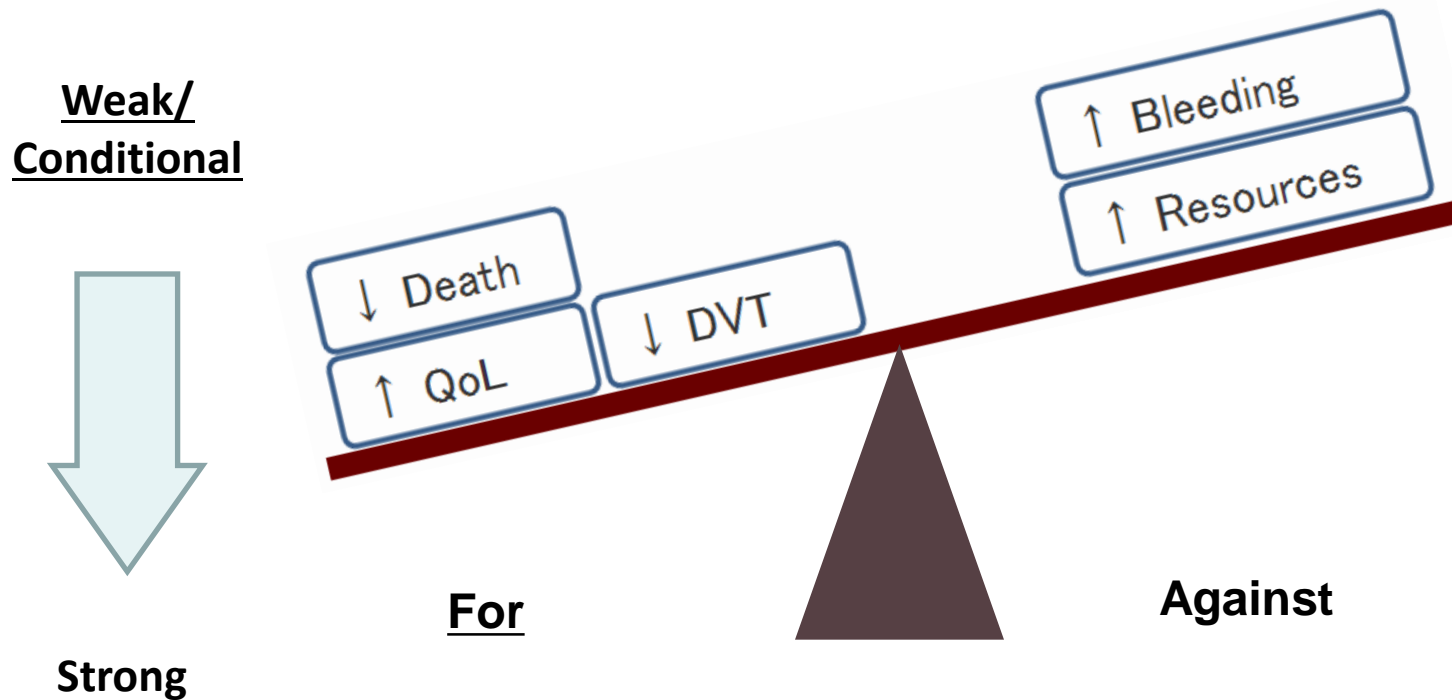
望ましい効果と望ましくない効果のバランス

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(図3.2.2-1)。



推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス



推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

NNT (number needed to treat), NNH (number needed to harm) を計算してみる。

- <http://homepage3.nifty.com/aihara/2x2.html>
- エクセル spreadsheet: http://homepage3.nifty.com/aihara/jama_evidence.html

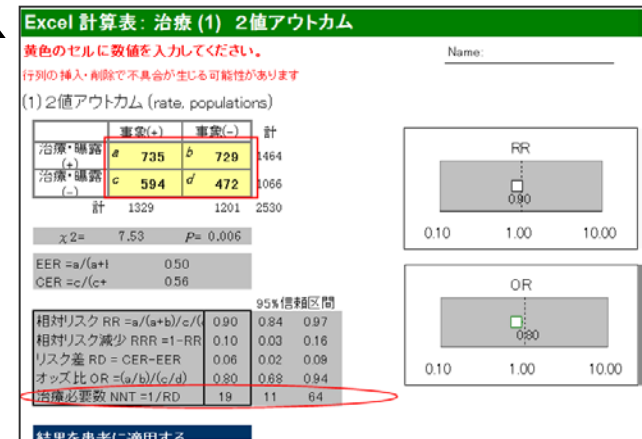
- 望ましい効果(利益)の大きさはどの程度だろうか？

アウトカム	NNT (95%CI)
全死亡率 (12mo)	18 (10~63)
症候性VTE	31 (20~61)

絶対リスク減少
0.055 (0.016~0.094)

- 望ましい効果(利益)の大きさはどの程度だろうか？

アウトカム	NNH (95%CI)
大出血	2532 (96~ -104)



推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

$$\text{LHH} = (1/\text{NNT}) / (1/\text{NNH})$$

LHH (Likelihood of Help vs Harm) という指標があるが、利益が害の140(=2532/18)倍であるとは安易に判断できない。

LHH: <http://www.grade-jpn.com/grade/guideline/LHH.pdf>

ベースラインリスクは？

推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

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。

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。

Profile: heparin versus no heparin for patients with cancer who have no other therapeutic or prophylactic indication					
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 ^{1,2,3} due to inconsistency	RR 0.93 (0.85 to 1.02)	557 per 1000	39 fewer per 1000 (from 84 fewer to 11 more)
				645 per 1000	45 fewer per 1000 (from 97 fewer to 13 more)
				791 per 1000	55 fewer per 1000 (from 119 fewer to 16 more)
Symptomatic VTE	2264 (7 studies) 12 months	⊕⊕⊕⊕ HIGH ¹	RR 0.55 (0.37 to 0.82)	62 per 1000	28 fewer per 1000 (from 11 fewer to 39 fewer)
				29 per 1000	13 fewer per 1000 (from 5 fewer to 18 fewer)
Major bleeding	2843 (9 studies) 12 months	⊕⊕⊕⊕ MODERATE ^{1,4} due to imprecision	RR 1.3 (0.59 to 2.88)	19 per 1000	6 more per 1000 (from 8 fewer to 35 more)
				7 per 1000	2 more per 1000 (from 3 fewer to 13 more)
Minor bleeding	2345 (7 studies) 12 weeks	⊕⊕⊕⊕ MODERATE ^{1,4}	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)

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。

推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

どのように判断すべきか

推奨作成者は、望ましい効果と望ましくない効果のバランスを検討する際に、いくつかの要因について検討すべきである。

- アウトカムの重要性
- アウトカムのベースラインリスク
- 介入の相対効果と絶対効果
- 効果推定値の精確さ
- コスト

<http://homepage3.nifty.com/aihara/GRADEproHelp/Balance.htm>

推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

●アウトカムの重要度：全死亡率と大出血、症候性VTEを重大(critical)と判定した。

●効果推定値の精確さを評価するにあたって、以下のように考えるべきである。

統合推定値が介入を支持するものの、その一方でCIが“効果なし”を含む場合、

「考えられる最大の効果を表すCIの境界が、真の潜在的効果を示していたとしたら、その介入を推奨するだろうか？」あるいは、

「考えられる最も悲観的な効果を表すCIの境界が真実を反映していたとしても、その介入を支持する推奨を作成するだろうか？」

●ガイドラインパネルがある介入を強く推奨する場合、それは「望ましい効果が望ましくない効果を十分に上回り、かつ正確な説明を受けたほぼすべての患者が治療を受けることを選択する」、とガイドラインパネルが確信をもつことを意味する。

推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

臨床シナリオ(CQ-1)の場合、

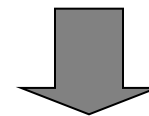
全死亡率として、12ヶ月の死亡率(RR 0.93: 0.85~1.02)と症候性VTE (RR 0.55: 0.37-0.82)、大出血(RR 1.3: 0.59~2.88)を比較する。

「考えられる最大の効果を表すCIの境界(この場合、NNT 10、NNT 20)が、真の潜在的効果を示していたとしたら、大出血のリスク(最悪として、NNH 96)と比較しても、介入を推奨するだろう」

「考えられる最も悲観的な効果を表すCIの境界(この場合、NNT 63, NNT 61)が真実を反映していたとしても、大出血のリスク(最悪として、NNH 96)と比較してもその介入を支持する推奨を作成するかもしれない」

「望ましい効果が望ましくない効果を十分に上回り、かつ正確な説明を受けた**ほぼすべての患者**が治療を受けるだろうかという点では、確信を持ってない」

アウトカム	NNT (95%CI)
全死亡率	18 (10~63) *High risk group
症候性VTE	31 (20~61)
アウトカム	NNH (95%CI)
大出血	2532 (96~ -104)



推奨する。推奨の程度は弱。

2 あるいは 1?

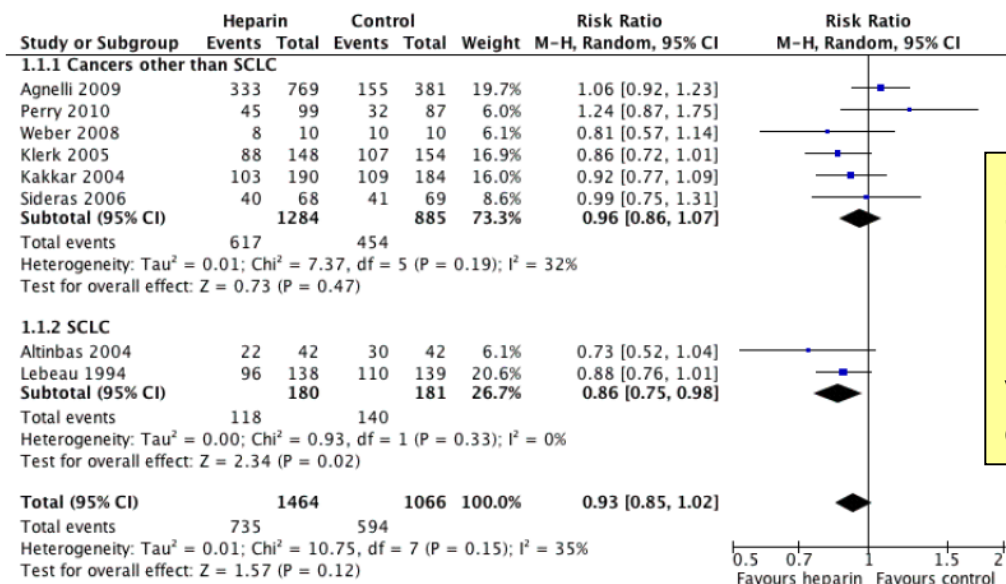
推奨の強さの要因 -②

望ましい効果と望ましくない効果のバランス

固形癌患者を一括して扱うことには抵抗がある、推奨を決定する場合には集団別(固形癌のタイプ別)に推奨の作成を検討すべきという意見が出るかもしれない。

サブグループではどうか？

Figure 3. Subgroup analysis comparing SCLC with other types of cancer for mortality at 12 months



サブグループ解析(小細胞肺癌と他のタイプの癌の比較)において、両者の間には統計的に有意差があった(p=0.03):
 RR 0.86 (95%CI 0.75 to 0.98) for SCLC versus RR 0.96 (95% CI 0.86 to 1.07) for other types of cancer.

推奨の強さの要因 -③ 価値観と好み

「価値観と好み」(Value and preference) の意味：

価値観と好みは、健康と生活に対する患者の視点、信条、期待、目標を総括する用語である。より正確には、この用語は管理選択肢について考えられる利益、害、コスト、限界、不便さを比較考量するために各個人が使用するプロセスを指す。「価値観」という用語がこれらのプロセスに最も近いニュアンスを持つとする意見と、「好み」というニュアンスが選択という概念を最も如実に捉えているとする意見があることから、われわれはこれらの用語をセットで使用することで、この概念を伝達している。

どの治療にも、必ず長所と短所がある。通常、患者は治療を選択するにあたり、疾患のアウトカム、治療の合併症、治療のための患者と家族の負担、治療に関するコストなどを把握し、その意味を理解し、自身の価値観や好み(value and preference)を基準に、その治療を受けるかどうかを判断する。患者個人の価値観や好みが確実でない(uncertainty)、あるいは価値観と好み患者間でばらつく場合は、推奨度を低くする。

診療ガイドラインのためのGRADEシステム、「3.2-3 価値観や好み」参照のこと。
具体的な記述例としては、同書 p125参考。

推奨の強さの要因 -③ 価値観と好み

臨床シナリオ例の場合、以下のように判断する。

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」

推奨の強さの要因 -④ コスト・資源利用

コストや資源の利用に関するエビデンスの質の評価は、他のアウトカムの評価とはやや異なる。

資源の利用を考慮する場合、ガイドラインパネルはコストについて検討する前に、まずその他のアウトカムに関するエビデンスの質を決定し、利益と不利益とを天秤にかけなければならない。もし健康への正味の利益に関する情報が欠如している場合は、資源の影響について考慮しても意味がない。また、介入の利益が不利益を大幅に上回る場合は、介入の実施の意思決定に際し資源の利用について検討する重要性は低くなる。コストは、利用する資源の金銭的価値であり、推奨において重要な検討項目である一方で、状況に依存し(context specific) 時間によって変化するために、コストの評価は困難なことも多い。

GRADE では、ガイドラインパネルが資源の利用について考慮する場合、資源の利用を考慮した推奨事項は1件のみの策定とするよう提案している。2件の推奨(資源の利用を考慮しない推奨を1件と、資源の利用を考慮した推奨を1件)を策定することは回避すべきである。

(注)

費用効果分析や費用効用分析などは意思決定の際の参考にはなるものの、これらの多くが限界やバイアスを含む可能性があり、GRADE は、通常のアウトカムについてのエビデンス・プロファイルにはこれらの分析を含めないことを推奨している。

推奨の強さの要因 -④ コスト・資源利用

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。

*High risk group

(**) 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。

本計算は、Cost of Preventing an Event (**COPE**) に準じた。

ACP J Club 2008 148; A-8: Adding cot to NNT: the COPE statistic

推奨の強さの判定基準と判定

推奨の強さの決定要因	判定	説明
<p>望ましい効果と望ましくない効果のバランスが不確実 (コストは考慮しない)</p> <p>正味の利益が少なく、望ましい効果と望ましくない効果のバランスについて不確実性が大きいほど、推奨度が「弱い」とされる可能性が高くなる。</p>	<p><input checked="" type="checkbox"/> はい</p> <p><input type="checkbox"/> いいえ</p>	<p>利益が不確実である。</p>
<p>エビデンスの質が低い</p> <p>全体的なエビデンスの質が低い(または非常に低い)ほど、推奨度が「弱い」とされる可能性が高くなる。</p>	<p><input type="checkbox"/> はい</p> <p><input checked="" type="checkbox"/> いいえ</p>	<p>エビデンスの質は「moderate」である。</p>
<p>患者の価値観や好みの不確実さ、あるいは相違</p> <p>価値観や好みにばらつきがあるほど、または価値観や好みにおける不確実性が大きいほど、推奨度が「弱い」とされる可能性が高くなる</p>	<p><input type="checkbox"/> はい</p> <p><input checked="" type="checkbox"/> いいえ</p>	<p>死亡リスクを考慮すると治療のマイナス面は少ないことからほとんどの患者が治療を受け入れると考えられる。</p>
<p>正味の利益がコストや資源に見合うかどうか不確実</p> <p>介入のコストが高いほど、すなわち消費される資源が多いほど、推奨度が「強い」とされる可能性が低くなる)</p>	<p><input type="checkbox"/> はい</p> <p><input checked="" type="checkbox"/> いいえ</p>	<p>薬剤のコストはそれほど高くない。</p>

* 「はい」が多いと、推奨度が「弱い」とされる可能性が高くなる。

推奨度のグレーディング

患者にとって重大なアウトカムに関する全体的なエビデンスの質は「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」

(GRADE 2B)

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」

- 推奨の強さ(推奨度)は2種類、1「強い」、2「弱い(条件付き)」
- エビデンスの質は4種類、A「高」、B「中」、C「低」、D「非常に低」
- ガイドライン利用者にとって優先すべき医療行為は推奨度であることから、GRADEでは、「推奨度」、「エビデンスの質」の順で記載する。

推奨作成 癌患者における非経口的抗凝固療法(ヘパリン)

疑問: 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
患者: 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)。
介入: 非経口ヘパリン
比較: ヘパリン投与なし
アウトカム: 全死亡率、大出血、小出血、症候性DVT、QOL、血小板減少。

エビデンスの要約:(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。

エビデンスの質:

全死亡率に関するエビデンスの質は、非一貫性(inconsistency: $I^2=41\%$ 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。

最良の推定値:

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。

利益と害、負担、コストの判定:

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。

推奨のグレーディング:

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。

GRADEpro を使った推奨判定

<http://ims.cochrane.org/revman/gradepro>

GRADEpro [anticoagulation for cancer patients.grd]

File Add View Options Help

New Open Print Save Undo all changes Add profile group Add profile Add outcome Import from RevMan Preview SoF table

Profiles tree << Edit

Parenteral anticoagulation in
heparin versus no heparin

- All-cause mortality at 12 weeks
- All cause mortality at 24 weeks
- Symptomatic VTE
- Major bleeding
- Minor bleeding
- Health related Quality of Life

Evidence profile

Question: Should heparin versus no heparin be used for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation?

SoF title: heparin compared to no heparin for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation

Format:

Intervention: heparin versus no heparin

Health problem: patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation

Setting: outpatient

Time frame:

Profile Information

Bibliography (systematic reviews):
Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation
Akl EA, S Gunukula, M Barba, VED Yosuiico, FF van Doormaal, S Kuipers, S Middeldorp, HO Dickinson, A Bryant, H Schunemann. Cochrane Database Syst Rev. 2012;(4):CD009754.
Profile author(s):
aihara//

Created on: 2012/04/11

Last major update: 2012/04/17

Delete Add Outcome **Go to Recommendation**

Profile: heparin versus no heparin for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation

Participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95%CI)	Anticipated absolute effects	
							With	With		Risk with	Risk difference with

GRADEpro ver.3.6では、
[Go to Recommendation]

GRADEpro を使った推奨判定

GRADEprofiler
version 3.0.1
Version: 3.6

Profiles tree << Edit

- Parenteral anticoagulation
 - heparin versus no heparin
 - All-cause mortality at 90 days
 - All cause mortality at 90 days
 - Symptomatic VTE
 - Major bleeding
 - Minor bleeding
 - Health related Quality of Life

Question: **Should heparin versus no heparin be used for patients with cancer who have no other therapeutic or prophylactic indic...**

Quality of evidence across all critical outcome: moderate ⊕⊕⊕○

Evidence profile

Absence of high quality evidence	yes	Explanation
Uncertain balance of benefits and harms/burden	yes	Explanation
Uncertain or different values and preferences	no	Explanation
Uncertain if net benefits are worth the costs	no	Explanation

Voting results

Strong for intervention	Weak for intervention	No specific recommendation	Weak against intervention	Strong against intervention
4	10	3	2	1

Strength of recommendation: weak

Draft recommendation: We suggest parenteral heparin for patients with cancer who have no other therapeutic or prophylactic indication.

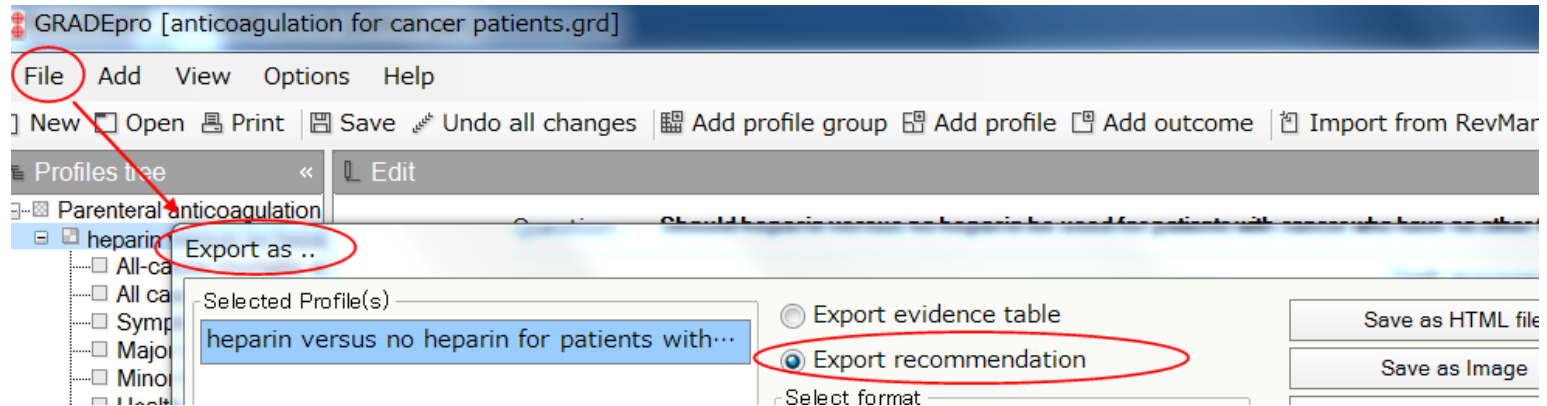
Values and preference statement: This recommendation places a relatively high value on death and thromboembolic events, and a relatively low value on

Remarks: Voting: 70% of panel is agreement with a recommendation (50% weak recommendation) , and 15% against

Go to Profile Information

推奨度の判定のための4要因について評価する

GRADEpro を使った推奨判定



Draft recommendation for consideration by the guideline panel

- Draft recommendation**
We suggest parenteral heparin for patients with cancer who have no other therapeutic or prophylactic indication.
- Values and preferences associated with this recommendation**
This recommendation places a relatively high value on death and thromboembolic events, and a relatively low value on bleeding by parenteral heparin.
- Remarks**
Voting: 70% of panel is agreement with a recommendation (50% weak recommendation) , and 15% against recommendation.

Overall quality of evidence across all critical outcomes: MODERATE

この推奨草案の詳細を、以下のリンクで公開。

[grade_grid_20120420](http://www.grade-jpn.com/grade_grid_20120420)

http://www.grade-jpn.com/grade_grid_20120420.html

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。^{1,2)}

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。

1st 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”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を活用する」という方針である。

RAND/UCLA を使った評価例*

シナリオ_0616において、1～9点スケール(1:同意できない、 9:同意できる)を使って同意の程度を入力してください。

(1) エビデンスの質評価 (moderate quality of evidence)について、
 (2) 推奨の強さと方向について (weak/conditional ↑?) について、1～9点スケールによる同意の程度を入力してください。

投票結果(1): http://groups.yahoo.co.jp/group/grade_meeting/surveys?id=1290842
 投票結果(2): http://groups.yahoo.co.jp/group/grade_meeting/surveys?id=1290843

rand.xls													
	A	B	C	D	E	F	G	H	I	J	K	L	
1													
2													
3	評価項目				評価点					評価点の	見解不一致指数		
4	全体的なエビデンスの質 (moderate)		1	2	3	4	5	6	7	8	9	median	0.17
5	投票の結果												
6	選択肢		票数	得票率	全投票数5票								
7	1: 全く同意できない		0	0.00%									
8	9: 完全に同意できる		1	20.00%									
9	2		0	0.00%									
10	3		1	20.00%									
	4		0	0.00%									
	5		0	0.00%									
	6		1	20.00%									
	7		2	40.00%									
	8		0	0.00%									
11													
12	推奨の強さと方向 (weak rec. for)				1		2	2				7.0	0.14
13	投票の結果												
14	選択肢		票数	得票率	全投票数5票								
15	1: 全く同意できない		0	0.00%									
16	9: 完全に同意できる		0	0.00%									
17	2		0	0.00%									
18	3		0	0.00%									
19	4		0	0.00%									
20	5		1	20.00%									
21	6		0	0.00%									
	7		2	40.00%									
	8		2	40.00%									

この調査と投票および解析は、6/16の会議前に実施したものである。

評価点の中央値、見解不一致指数から、いずれも “appropriate” と判定できる。

合意形成のためのGRADE grid

GRADE gridとは、ガイドライン作成においてパネル全員からの意見を収集するための、迅速かつ明確、客観的な統計学に基づく手法である。質の高いGRADE evidence profileを作成するには多大な資源を要するが、Gridを使うことは、判断の迅速性、透明性、効率性の向上、コスト負担の回避につながる。

If consensus is not reached by discussion, this table below allows the panel making a recommendation to record their views (votes) about the recommendation related to a specific interventions, based on their analysis of the available evidence, the benefits and downsides, values and preferences and cost. This assessment is then mapped to the strength of recommendation for the use, or non-use, of each intervention.

Insert the number of votes for the recommendation in each category

GRADE	Strong	Weak	No strength	Weak	Strong
Assessors view of the balance of desirable and undesirable consequences of the intervention	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
Recommendation	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

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。

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

GRADEpro を使った推奨判定

Recommendation (final)

We **suggest** parenteral heparin for patients with cancer who have no other therapeutic or prophylactic indication.

(weak recommendation, moderate quality evidence)



http://www.grade-jpn.com/grade_grid_20120420.html を参照



GRADE ワークショップ資料(治療介入)

GRADEワークショップのための資料として、役立つものをweb上に公開しているので、本資料とあわせて活用していただきたい。システムティックレビューのエビデンスの質評価から推奨の強さを判定するまでのプロセスを理解できるはずである。

・GRADEワークショップ資料(2008) (*)

http://homepage3.nifty.com/aihara/GRADE-workshop_2008_j.pdf

このワークショップ資料は、Schunemann教授らが作成した資料を翻訳したものであり、書籍「診療ガイドラインのためのGRADEシステム」の4.3章と同じものである (p134-143)。

Web上でも、公開している。



http://www.grade-jpn.com/grade_materials.html

Part 3

GRADEの長所・限界と使用基準

GRADE's advantages and limitations

Criteria for using GRADE

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、患者が利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。
2. GRADEの“長所”と“限界”を認識する。
3. GRADEワーキンググループが作成した、「GRADEを利用したとする基準」を参考にする。



GRADEの長所

他システムと比較したGRADE の長所

1. ガイドライン開発を代表する人たちが国際的に共同して作成や評価をしている。
2. エビデンスの質と推奨度を明確に分離している。
3. “患者にとって重要なアウトカム”について検討している。
4. エビデンスの質の評価基準を明確にしている。
5. エビデンスから推奨への移行プロセスを明確にしている。
6. 推奨を作成する際に患者の価値観や好みを考慮することを明確にしている。
7. 臨床医、患者および政策立案者に対して、推奨度の解釈を明確にしている。
8. システマティック・レビュー、医療技術評価および診療ガイドラインに役立つ。



GRADEの限界

GRADEには、5つの重要な限界がある。

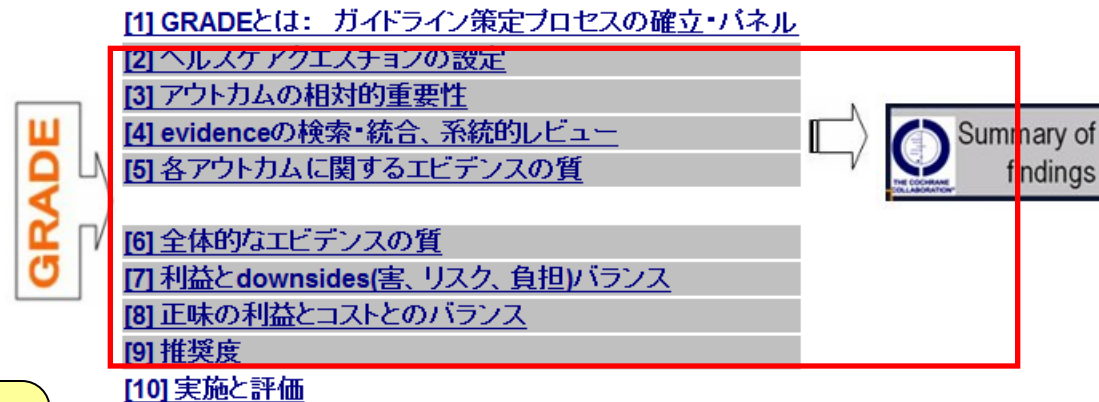
1. 現状ではGRADEは、介入または政策の比較という2つの管理戦略に関する疑問に対応するために開発されたものであり、予後や病因、あるいはリスクに関する疑問を対象としてはいない。
2. 大義名分的な、不明瞭に定義された一連の推奨事項にGRADEを適用しようとするのには問題がある。議論の余地がないような推奨、賢明な人であれば反対事項を支持しようとは考えもしない推奨や、推奨事項が曖昧すぎてエビデンスを問題に適用できないような部分には、GRADEの適用対象外とすべきである。
3. GRADEは、診断検査に適用されるよりは、予防や治療介入に適用されることが多く、公衆衛生や医療システムに関する疑問よりは、臨床的な疑問に適用されるケースが多い点である。特に診断研究にGRADEを適用する場合は、エビデンスの質判定基準の一部を改変する必要性が生じる。
4. システマティック・レビューや診療ガイドラインの作成では、GRADEを適用する前後に、いくつかの基本的かつ一般的なプロセスが存在していることである。このため、プロセス全体の中でGRADEがどの部分に適用されるのかを理解し、GRADE適用対象外のステップについてはGRADE以外に手引きを求める必要がある。
5. GRADEは、系統的かつ明示的な、エビデンスの質および推奨度の評価を可能にするが、エビデンスの質判定の不一致を解消することを目的としたものではない。すなわち、同じエビデンスでもエビデンスの質の判定が異なることがしばしばあるが、GRADEで確実にできることは、その見解の不一致の本質的な理由を把握することである。



GRADEの適用範囲

GRADE

GRADEシステムとGRADEガイドライン作成



ステップ [2]から[9]までが
GRADE適用範囲である

GRADEシステムとGRADEガイドライン作成

http://www.grade-jpn.com/grade/guideline/grade-guideline_2008-nov.html

GRADEを使用したという基準

GRADEシステムを利用したと言えるための基準

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（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、結核出版バイアス、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（a）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）、一般的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。

本内容は、**GRADE working groupの今後の正式な発表論文をもとに改変する予定です。**

結 論

- 診療ガイドラインは、入手可能な**最良のエビデンス**に基づくべきである。
- 意思決定や判断において透明性が鍵となる。
- **GRADE アプローチ**
 - simple, transparent, systematic



あなたは、GRADEシステムの使い方をほぼ理解できたと思います。

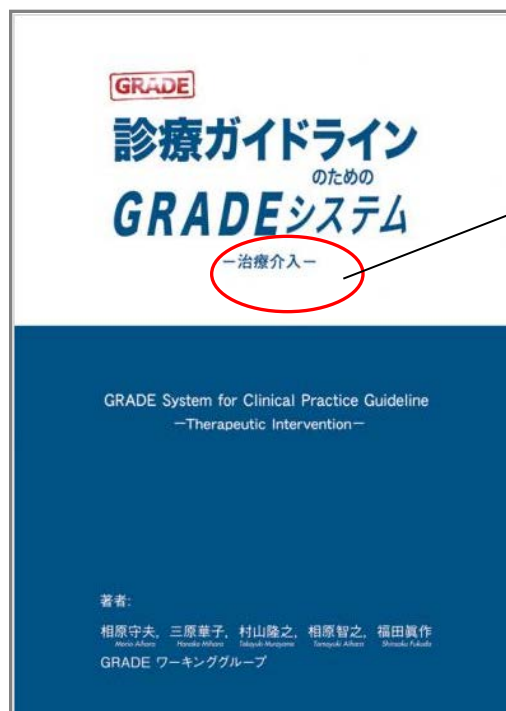
・・・では次は？



診療ガイドラインのためのGRADEシステム

(相原守夫、他、凸版メディア出版、2010)

<http://homepage3.nifty.com/aihara/grade-cover.html>



GRADEに関する国内唯一の教科書であろう(*)

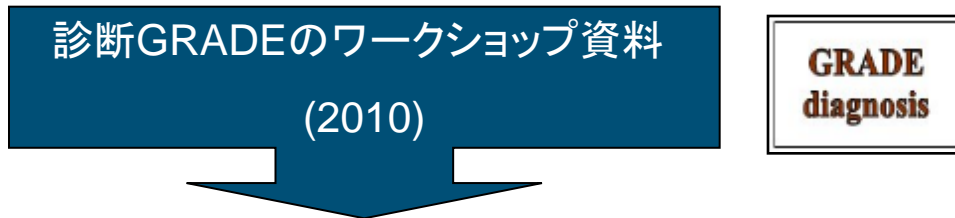
— 治療介入 —

GRADEの基本は、
“治療介入GRADE” であるが...



診断研究のGRADE

診断検査・診断戦略を扱うGRADEは、治療介入のGRADEと基本的には同じであるが、検査精度は患者にとって重要なアウトカムの代理アウトカムである。診断検査・診断戦略のエビデンスの質の判断は治療介入とは異なる点がある。



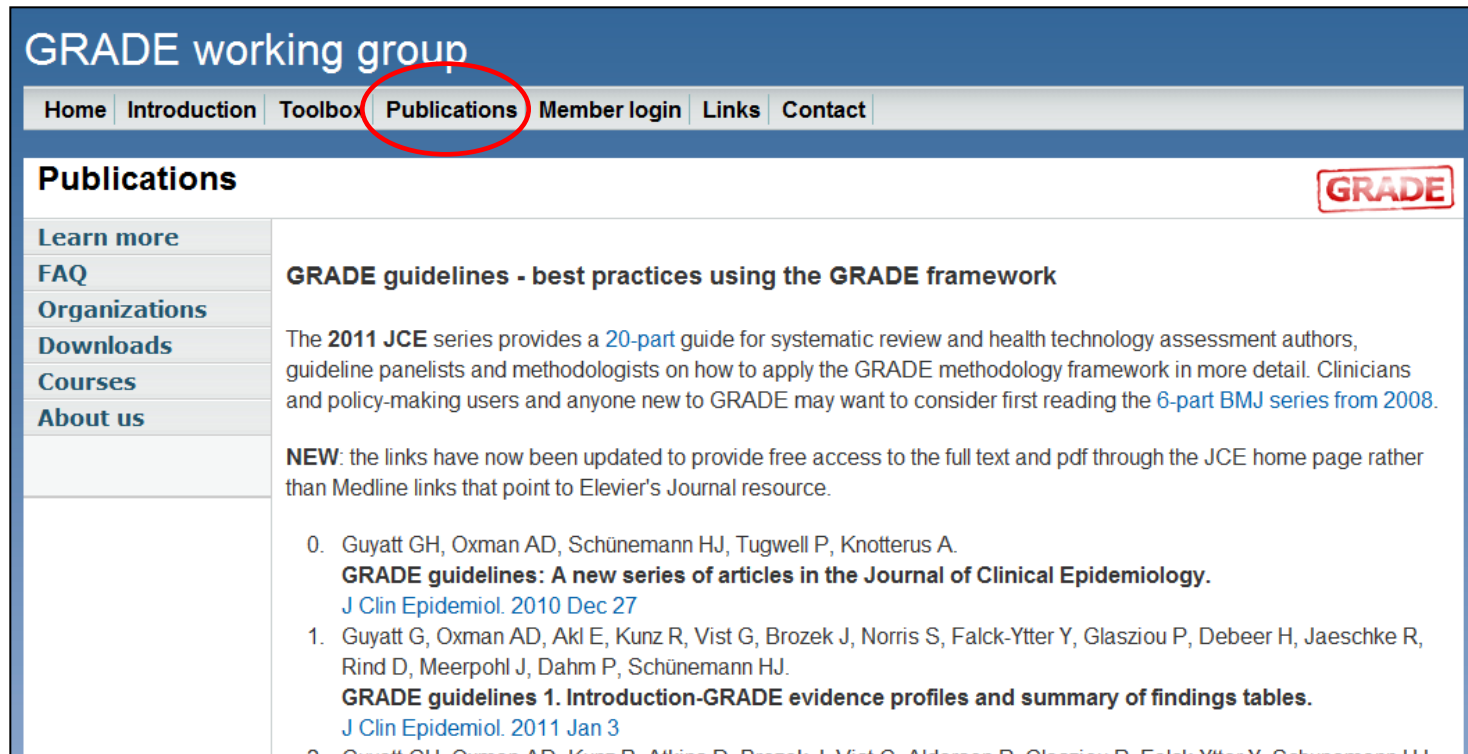
http://homepage3.nifty.com/aihara/GRADE-Dx_workpackage_20101016_j.pdf

診断GRADEに関する日本語資料としては、(現時点では)国内には他に見当たらない。



GRADEの今後

GRADE working group: Publications



The screenshot shows the website for the GRADE working group. The navigation menu includes Home, Introduction, Toolbox, Publications (highlighted with a red circle), Member login, Links, and Contact. The main content area is titled "Publications" and features a sidebar with links to Learn more, FAQ, Organizations, Downloads, Courses, and About us. The main text discusses the 2011 JCE series and provides a list of publications.

GRADE

Publications

Learn more

FAQ

Organizations

Downloads

Courses

About us

GRADE guidelines - best practices using the GRADE framework

The **2011 JCE** series provides a [20-part](#) guide for systematic review and health technology assessment authors, guideline panelists and methodologists on how to apply the GRADE methodology framework in more detail. Clinicians and policy-making users and anyone new to GRADE may want to consider first reading the [6-part BMJ series from 2008](#).

NEW: the links have now been updated to provide free access to the full text and pdf through the JCE home page rather than Medline links that point to Elsevier's Journal resource.

0. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A.
GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology.
[J Clin Epidemiol. 2010 Dec 27](#)
1. Guyatt G, Oxman AD, Akl E, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ.
GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables.
[J Clin Epidemiol. 2011 Jan 3](#)
2. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson D, Glasziou P, Falck-Ytter Y, Schünemann HJ

GRADEの今後

GRADE working group: Organizations

GRADE working group






Home Introduction Toolbox Publications Member login Links Contact

Organizations

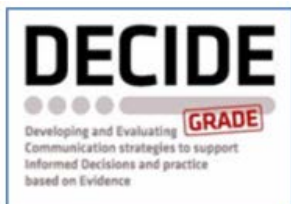
Learn more
FAQ
Organizations
Downloads
Courses
About us

Organizations that have endorsed or that are using GRADE*

Some organizations in their enthusiasm to use GRADE, have modified the GRADE approach. We recommend against such modifications because the elements of the GRADE process are interlinked, because modifications may confuse some evidence and guideline users, and because such changes compromise the goal of a single system with which clinicians, policy-makers and patients can become familiar.

	World Health Organization - International	<i>Example</i>
	Endocrine Society - USA	<i>Example</i>
	American College of Chest Physicians - USA	<i>Example</i>
	UpToDate - Putting Clinical Information Into Practice - USA	<i>UTD GRADE tutorial</i>
	Agenzia sanitaria regionale, Bologna - Italia	

GRADEの今後



DECIDE (Developing and Evaluating Communication Strategies to Support Informed Decisions and Practices Based on Evidence)

<http://www.decide-collaboration.eu/welcome>

DECIDE プロジェクトを翻訳したものの一部を紹介。

DECIDE研究プロジェクトの概要

計画目標:

「対象を絞ったガイドライン配布のための方法を開発・評価するために、GRADE Working Groupの取り組みを土台とし、エビデンスに基づく推奨の配布を改善する」

方法:

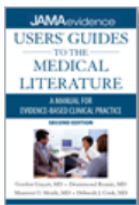
GRADE Working Groupのメンバーにより結成されるDECIDE協会 (DECIDE consortium) は、このアプローチをさらに発展させることで、診療の行方を左右する主な利害関係者 (医療専門家、政策決定者と管理責任者、患者と一般市民) に対し、エビデンスに基づく推奨を効果的に配布することを確実にする。われわれは、諮問グループ、審議、ユーザー調査を通じて利害関係者からのインプットを集める。

期待される結果:

多様なセッティングで厳密に評価された推奨配布戦略は、研究の診療への反映を支援し、実際の医療システムに適用される。

<http://homepage3.nifty.com/aihara/decide.doc> より



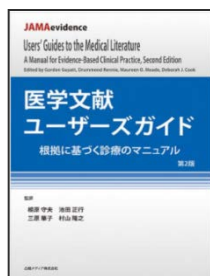


Users' Guides to the Medical Literature: A Manual for Evidence- Based Clinical Practice, 2nd Edition

Gordon Guyatt, Drummond Rennie, Maureen O. Meade,
and Deborah J. Cook



翻訳



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 [単行本]

相原守夫 (著), 池田正行 (著), 三原華子 (著), 村山
隆之 (著)

http://homepage3.nifty.com/aihara/jama_evidence.html

GRADEの基本は、JAMA Users' Guide to the Medical Literatureに記載されているEBMであり、本書を日本語翻訳したものが「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」である。