

追加資料-⑩：GRADEワークショップ資料：治療介入（RCTと観察研究）

このワークショップ資料は、GRADE ワーキンググループが作成したもの（GRADE intervention workshop package v20110822）を改変したもので、GRADEpro もしくは G²DT を利用できる環境で活用するのが望ましいが、本資料に含まれるワークシートをそのままコピーして利用することで GRADE の理解に役立てることも可能である。

● 本資料に含まれるワークシートと表

資料	資料名	頁
ワークシート 1	意思決定のためのアウトカムのリスト	6
ワークシート 2	各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評価	7
ワークシート 3	エビデンスプロファイル	8
ワークシート 4	推奨のための決断テーブル	9, 10
表 1	エビデンスの質評価基準と定義	11
表 2	研究の限界：RCT	12
表 3	研究の限界：観察研究	13, 14
表 4	risk of bias 以外のグレードダウン要因に関する判定	15
表 5	グレードアップ要因に関する判定	15
表 6	標準様式のエビデンスプロファイル例	16, 17
表 7	ACCP 様式のエビデンスプロファイル例	18, 19
表 8	SoF テーブル例	20, 21
表 9	Evidence to Decision テーブル例	22, 23

● ワークショップにおける課題

- 小グループで作業を行う。
- グループ全体への報告を行う担当者を選ぶ（メモをとる）。
- 時間を管理する。
- 次頁の指示に従う。

課題の途中で、診断に関するシステマティック・レビューの SoF テーブルまたはエビデンスプロファイルがどのように提示されるか、または最終的に提示するものがどれほど改善されるかを考えていただきたい。

● 指示 (要点：指示の詳細は各セクションをみるとよい)

1. システマティック・レビューの詳細と提起されたクリニカル・クエスチョンを特定する
レビュー (添付) の抄録を読み、クリニカル・クエスチョンとその構成成分を PICO (患者集団, 介入, 比較, アウトカム) 形式で定式化し、主な比較を同定する (本資料の 3 頁目)。
2. 意思決定に重要なアウトカムを選択する
患者にとって重要なアウトカムを選択する (ワークシート 1)。
3. アウトカムの相対的な重要性を評価する
患者にとって重要なアウトカムを最大 7 個まで選択し、各アウトカムの相対的な重要性を決定する (ワークシート 1, ワークシート 3)。
4. 各アウトカム (each outcome) に関するエビデンスの質を評価する
各アウトカムに関するエビデンスの質評価をワークシート 2, ワークシート 3 に記入する。
エビデンスの質評価には、表 1 (エビデンスの質評価基準と定義)、表 2 (研究の限界：RCT)、表 3 (研究の限界：観察研究)、表 4 (risk of bias 以外のグレードダウン要因に関する判定)、表 5 (グレードアップ要因に関する判定) (11 頁目～15 頁目) を参考とする。

注：ワークシートを使わず、GRADEpro や G²DT を使う場合は、各アウトカムのエビデンスの質が自動的に表示される。
5. 望ましいアウトカムと望ましくないアウトカムに関する、効果の大きさ (Summary of Finding: SoF) に関する情報を提示する
エビデンスプロファイル (ワークシート 3) を完成させる。
6. 全ての重大なアウトカムにわたる全体的なエビデンスの質を等級付けする
完成されたエビデンスプロファイル (ワークシート 3)、または GRADEpro/G²DT により作成されたエビデンステーブルを利用して、アウトカム全般にわたる全体的なエビデンスの質を決定する。
7. 決断テーブルを使用してエビデンスから推奨を決定する
決断テーブル (Evidence to Recommendation テーブルまたは、Evidence to Decision テーブル) を利用して、最終的な推奨を決定する (ワークシート 4)。
8. 合意形成
もし、討議によって合意に至らなかった場合、ある特定の介入に関する推奨の強さに関する自らの見解 (投票) を記録することができる。また、各カテゴリの推奨に対する投票数を入力することもで

きる (ワークシート4)。

1. システマティック・レビューの詳細と提起されたクリニカル・クエスチョンを特定する

システマティック・レビューのタイトル：
対応している疑問を考えてみる。主な比較を選択する (例, ワルファリン vs アスピリン)。

1.1 クリニカル・クエスチョンを特定する。

患者 (P: 集団)	
介入 (I)	
比較 (C)	
アウトカム (O)	<p>[治療の有無にかかわらず、患者が経験しうるアウトカムを全て考慮する (ワークシート1を使用)]</p> <ul style="list-style-type: none"> ・ ・ ・ ・

1.2 最終的なクリニカル・クエスチョン

--

例：最近増悪をきたした慢性閉塞性肺疾患の患者に対しては、通常のコミュニティケアよりも呼吸リハビリテーションを実施すべきか。

2. 意思決定に重要なアウトカムを選択する

関連するアウトカムのリストを作成する ([ワークシート1](#) を参照)。

- 1次アウトカム (システマティック・レビューの著者が1次アウトカムであると位置づけたアウトカム) をリストアップする (コクランレビューの場合, 「Criteria for Considering Studies」の項を参照)。
- 報告されたアウトカム (別のアウトカムについてデータが報告されている場合は, そのアウトカム) を追加する (コクランレビューの場合, 「Results and Abstract」の項を参照)。
- レビューで報告されていないその他のアウトカムで, 重要であると考えられるアウトカムがあれば, 追加する (利益と有害作用の両方を含め, また関連があればコストも含める)。

3. アウトカムの相対的重要性を評価する

アウトカムのリストの中から, ガイドラインパネルや推奨作成者にとって最も重要で, エビデンスプロファイルに含めるべきであると考えられるアウトカムを選択する。

- 重要なアウトカムは最多で7件まで選択する。
- 選択したアウトカムを空のエビデンスプロファイル ([ワークシート3](#)) に転記するか, GRADEpro または G²DT を使って入力する。
- 小グループまたはペアを組んで作業すること。各小グループ内で, どのアウトカムが重要でエビデンスプロファイルに含めるべきか, について合意すること ([ワークシート1](#))。

4. 各アウトカムに関するエビデンスの質を評価する

- 当該アウトカムに関するエビデンスの質を決定するため, [ワークシート2](#), または[ワークシート3](#) に記入するか, GRADEpro または G²DT を使って入力する。
- エビデンスの質評価には, [表1 \(エビデンスの質評価基準と定義\)](#), [表2 \(研究の限界: RCT\)](#), [表3 \(研究の限界: 観察研究\)](#), [表4 \(risk of bias 以外のグレードダウン要因に関する判定\)](#), [表5 \(グレードアップ要因に関する判定\) \(11 頁目~ 15 頁目\)](#) を参考とする。

4.1 エビデンスの質の評価は, 研究デザインと研究の限界 (risk of bias) の評価から始める

- GRADE アプローチを使ったエビデンスの質評価は, 研究デザインから開始する。
- 次に研究の限界 (risk of bias) を含めたグレードダウンの5要因する。
- グレードダウンのない研究 (通常, 観察研究) について, グレードアップの3要因を考慮する。
- GRADE によるエビデンスの質は, 4段階 (高, 中, 低, 非常に低) のうちいずれか1つである。

5. 望ましいアウトカムと望ましくないアウトカムに関する, 効果の大きさ (Summary of Finding: SoF) に関する情報を提示する

- エビデンスプロファイル ([ワークシート3](#)) の “結果の要約 (Summary of Findings: SoF)” に

- 関するデータを入力して、エビデンスプロファイルを完成する（表 6, 表 7, 表 8 を参照）。
- 選択した全てのアウトカムに関して、介入効果の大きさに関する情報を、SoF 部に入力する。
 - GRADEpro または G²DT を使うと、エビデンスプロファイルや SoF テーブル、さらには後述の決断テーブルを簡単に作成できる。
 - 介入効果の入力の際には、ベースラインリスクを最多で 3 種類まで設定できる。入力した統合推定の相対効果と信頼区間から絶対効果として 1,000 人あたりのイベントが自動的に表示される。
 - わからないことがある場合、GRADEpro または G²DT のヘルプファイル（GRADE ハンドブック）を参照すること。GRADEpro の旧バージョンの日本語版は web 上に公開されており、G²DT についても日本語版を利用できる^a。
 - ノート型パソコンを利用してグループでデータ入力を行ったり、2 人 1 組に分かれて各組に 1 つまたは複数のアウトカムを割りあてたりしてもよい。
 - エビデンステーブルは WORD, HTML, PDF などの様式でエクスポートすることができる。

6. 全ての重大なアウトカムにわたる全体的なエビデンスの質を等級付けする

（すべてのアウトカムについてエビデンスの質を評価しているならば）

GRADE のルールにしたがって、アウトカム全般にわたる全体的なエビデンスの質を、一つだけ決定する。アウトカム全般にわたる全体的なエビデンスの質の決定ルールは、4.5-2.1 章を参照のこと。

7. 決断テーブルを使用してエビデンスから推奨を決定する

- 推奨のための決断テーブル（[ワークシート 4](#)）を利用して、推奨の強さを「強い」または「弱い / 条件付き」に判定する。
- G²DT による Evidence to Decision (EtD) テーブルの利用が効率的である（表 9 を参照）。

8. 合意形成

アウトカム全般にわたる全体的なエビデンスの質、利益と不利益のバランス、価値観と好み、そしてコストの分析結果に基づき、ある特定の介入に関する推奨の強さについて、合意を形成する。合意形成プロセスを明確にし、合意の基準を事前に設定することが重要である。合意形成には、統計学的手法としての、修正デルファイ（RAND 法）を利用できる（4.5-5 章を参照のこと）。

^a G²DT の使い方については、第 4.4-6 章を参照のこと。

● ワークシート 1：意思決定のためのアウトカムのリスト

意思決定にとって最も重要なアウトカム（その介入を使用するかしないかという意思決定にとって重要かもしれないアウトカム）を選択する。以下を考慮する。

システムティック・レビュー のタイトル			
<p>エビデンスプロファイルに含むべき最も重要なアウトカムを7個まで選択する。 これを空のエビデンスプロファイルのシート（本資料のワークシート3）に転記する。 各アウトカムについての相対的重要性を、1点（重要でない）～9点（重大）で等級付けする。</p> <ul style="list-style-type: none"> ● 1～3点：重要でない（エビデンスプロファイルに含めない）。 ● 4～6点：重要だが、意思決定には重大でない（エビデンスプロファイルに含めるべきかどうかは、他の重要なアウトカムとの兼ね合いである）。 ● 7～9点：意思決定にとって重大である（エビデンスプロファイルに含める）。 			
アウトカム	重要性	エビデンスプロファイル に含めるか	
1.		はい	いいえ
2.		はい	いいえ
3.		はい	いいえ
4.		はい	いいえ
5.		はい	いいえ
6.		はい	いいえ
7.		はい	いいえ
8.		はい	いいえ
9.		はい	いいえ
10.		はい	いいえ
11.		はい	いいえ
12.		はい	いいえ

● ワークシート 2：各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評価

(もし、GRADEpro もしくは G²DT を利用しているなら、このワークシートはとばす)

(表 1. 「エビデンスの質評価基準と定義」を参考)

エビデンスの質判定基準	グレーディング (各項目について、いずれかの判定を 1 つ選択)	コメント (質を下げる理由を説明)	エビデンスの質 (アウトカムごとに 1 つ選択)
第 1 アウトカム			
研究の限界	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕⊕⊕ 高 <input type="checkbox"/> ⊕⊕⊕○ 中 <input type="checkbox"/> ⊕⊕○○ 低 <input type="checkbox"/> ⊕○○○ 非常に低
結果の非一貫性	なし 深刻 (-1) 非常に深刻 (-2)		
エビデンスの非直接性	なし 深刻 (-1) 非常に深刻 (-2)		
データの不精確さ	なし 深刻 (-1) 非常に深刻 (-2)		
出版バイアス	なさそう ありそう (-1) 非常にありそう (-2)		
第 2 アウトカム			
研究の限界	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕⊕⊕ 高 <input type="checkbox"/> ⊕⊕⊕○ 中 <input type="checkbox"/> ⊕⊕○○ 低 <input type="checkbox"/> ⊕○○○ 非常に低
結果の非一貫性	なし 深刻 (-1) 非常に深刻 (-2)		
エビデンスの非直接性	なし 深刻 (-1) 非常に深刻 (-2)		
データの不精確さ	なし 深刻 (-1) 非常に深刻 (-2)		
出版バイアス	なさそう ありそう (-1) 非常にありそう (-2)		

(本来、最多で第 1 ～ 7 アウトカムまでのチェックリスト表があるが、紙面の都合で第 3 アウトカム以後は割愛)

● ワークシート 3：エビデンスプロファイル

著者： _____ 月日： _____
 疑問： _____ セッティング： _____
 論文： _____

研究数	質の評価 (Quality assessment)					結果の要約 (Summary of findings: SoF)				質	重要性	
	デザイン	限界	非一貫性	非直接性	不精確さ	その他の要因	参加者数	対照	相対 (95%CI)			効果
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												
アウトカム 6												
アウトカム 7												
例：癌の化学療法による完全奏効率												
5	RCT	深刻な限界なし	深刻な非一貫性なし	深刻な非直接性なし	深刻な不精確さなし	出版バイアスあり	216/344	211/344	RR 1.0 (0.92 ~ 1.1)	-0/1000 (-49 ~ 61)	⊕⊕⊕ 中	重大

● ワークシート 4：推奨のための決断テーブル

推奨草案		
推奨に関連する価値観と好み（検討した各アウトカム別に、一連の価値観を想定する）		
(例： この推奨の作成に当たっては、致死率の高い疾患における死亡の回避を重要視した)		
重大なアウトカム全般にわたる全体的なエビデンスの質		
推奨の強さの判定（以下の4項目について判定し、その説明を記載する）		
推奨の強さに関する要因	判定	説明
高いまたは中等度の質のエビデンス (「高」または「中」の質のエビデンスはあるか?) 全体的なエビデンスの質が高いほど、強い推奨の可能性が高くなる。 エビデンスの質が低いほど、条件付き / 弱い推奨の可能性が高くなる。	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
利益と害・負担のバランスに関する確実性 (利益が害や負担を上回るあるいは下回ることに確信があるか?) 望ましい効果と望ましくない効果の差が大きければ大きいほど、推奨の強さが強くなる可能性が高い。正味の利益が大きければ大きいほど、強い推奨の可能性が高くなる	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
価値観と好み (想定された価値観について確信があるか、また標的集団において類似しているか?) 価値観と好みにばらつきが少ないほど、または確実性が大きいほど、強い推奨の可能性が高くなる。	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
資源の影響 (消費される資源は期待される利益に見合うか?) 介入のコストが低ければ低いほど、すなわち消費される資源が少ないほど、強い推奨の可能性が高くなる	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	

「はい」の回答が多いと、推奨の強さが「強い」とされる可能性が高くなる。

討議によって合意に至らなかった場合、推奨を策定するパネルは、以下の表を活用することによって、入手可能なエビデンス、利益と不利益、価値観と好み、そしてコストの分析結果に基づき、ある特定の介入に関する推奨についての自らの見解（投票）を記録することができる。この評価結果は各介入の使用・非使用を示す推奨の強さにマッピングされる。

次表の各カテゴリの推奨に対する投票数を入力すること。

各カテゴリの推奨の投票数を記入する				
介入の望ましい効果と望ましくない効果に関する評価者の見解	望ましい効果が望ましくない効果を明らかに上回る	望ましい効果が望ましくない効果をおそらく上回る	望ましくない効果が望ましい効果をおそらく上回る	望ましくない効果が望ましい効果を明らかに上回る
推奨の強さ	強い推奨	条件付き（弱い）推奨	条件付き（弱い）推奨反対	強い推奨反対
推奨の表現	“実施する”ことを推奨する	“実施する”ことを提案する	“実施しない”ことを提案する	“実施しない”ことを推奨する
投票数				

推奨の強さ	<input type="checkbox"/> 強い推奨 <input type="checkbox"/> 条件付き（弱い）推奨
-------	--

最終的な推奨（全パネルメンバーからの意見をまとめ最終的に決定する）	
推奨の強さ：	エビデンスの質：
根底にある価値観と好みについての想定	
注意	

表 1 GRADE エビデンスの質評価基準と定義

研究デザイン	body of evidence の初期評価	等級を下げる* (グレードダウン)	等級を上げる* (グレードアップ)	body of evidence の最終評価 (GRADE)
ランダム化試験	高 	<ul style="list-style-type: none"> • 限界 (risk of bias) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 • 非一貫性 (inconsistency) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 	<ul style="list-style-type: none"> • 効果の程度が大きい (large magnitude of effect) <ul style="list-style-type: none"> +1 大きな効果 RR>2 あるいは<0.5 +2 極めて大きい効果 RR>5 あるいは<0.2 	<ul style="list-style-type: none"> • 高: ⊕⊕⊕⊕ • 中: ⊕⊕⊕○ • 低: ⊕⊕○○ • 非常に低: ⊕○○○
観察研究	低 	<ul style="list-style-type: none"> • 非直接性 (indirectness) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 • 不精確 (imprecision) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 • 出版バイアス (publication bias) <ul style="list-style-type: none"> -1 ありそう -2 非常にありそう 	<ul style="list-style-type: none"> • 用量反応勾配 (dose-dependent gradient) <ul style="list-style-type: none"> +1 あり • すべての交絡因子 (plausible confounder) <ul style="list-style-type: none"> +1 提示された効果を減弱させている +1 効果が観察されていないのに当該効果を増大させる方向に働いている 	
* 1 = 1 段階グレードを下げる, あるいは上げる (例えば, 「高」から「中」へ) 2 = 2 段階グレードを下げる, あるいは上げる (例えば, 「高」から「低」へ)				
各アウトカムについての研究全般に関する body of evidence の質の定義				
<ul style="list-style-type: none"> • 「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。 • 「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, 推定値が変わる可能性がある。 • 「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, 推定値が変わる可能性が高い。 • 「非常に低」 = あらゆる効果推定値が不確実である。 				

例: あるアウトカムに関する複数の RCTs からのシステマティック・レビューを想定すると, 初めエビデンスの質は「高」から開始する。もし深刻な限界 (-1), さらに非常に深刻な不精確さ (-2) がある場合には, 最終的なエビデンスの質は, 「非常に低」となる。エビデンスの質評価は, 単純な加算ではないことに注意すべきである。

表2 研究の限界 (RCT)

RCT の限界 (risk of bias)	判定 (はい/不明/いいえ)
<ul style="list-style-type: none"> • 割り付けの隠蔽なし 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 	
<ul style="list-style-type: none"> • 盲検化の欠如 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。 	
<ul style="list-style-type: none"> • 患者やアウトカムイベントの不完全な検討 脱落率が 20% で、ITT 解析の原則を掲げながらも、その原則を遵守していない。 	
<ul style="list-style-type: none"> • 選択的アウトカム報告バイアス 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（典型的には、効果がないアウトカムのことが多い）。 	
<ul style="list-style-type: none"> • その他の限界 <ul style="list-style-type: none"> • 利益があったとして試験を早期中止する。 • 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効果がある^b。 • 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^c。 	

^b クロスオーバー試験は同一患者群が当該の2つ（あるいは全て）の治療を、時期をずらして受ける。つまり、介入も対照も、同一個人が時間をずらして体験する。対照者は同一個人のため、時間変化以外の潜在的な交絡因子を排除しやすい。対照は同一個人なので必要症例数を少なくできるが、持ち越し効果やいくつかの重大な欠点がある。

^c クラスター（cluster: 集団）単位で割り付けを行う方法で、病院単位や市などの単位といった集団レベルでランダム割り付けをする研究（通常の RCT では、多施設研究でも、個人レベルでランダム割り付けされる）。たとえば、60 の工場を、工場単位でランダムに治療群、対照群の半々に割り付けするような場合である。

表3 研究の限界 (観察研究)

観察研究の限界 (risk of bias)	判定 (はい/不明/いいえ)
<ul style="list-style-type: none"> • 曝露したコホートと曝露していないコホートが同じ集団から選択されていない <ul style="list-style-type: none"> • risk of bias 「低」 の例：曝露群と非曝露群が、同じ施設で同じ期間にわたって治療を受けた患者のデータが登録された同一の管理データベースから抽出されている。 • risk of bias 「高」 の例：曝露群と非曝露群が、異なる施設で、または異なる期間にわたって治療を受けている。 	
<ul style="list-style-type: none"> • 曝露の評価に確信が持てない <ul style="list-style-type: none"> • risk of bias 「低」 の例：確実な記録（例：外科手術の記録、薬局の記録）がある。複数回に及ぶ面接、またはその他の方法により、使用 / 曝露の現状を確認している。曝露群と非曝露群が、同じ施設で同じ期間にわたって治療を受けた患者のデータが登録された同一の管理データベースから抽出されている。 • risk of bias 「中」 の例：ある1時点で構造的面接を実施している。書面の自己報告である。過去の曝露の状態を振り返って確認するように求められた個人が、想起バイアスの影響を受ける可能性がある（有害アウトカムを発症していない場合は曝露を想起する可能性が低く、有害アウトカムを発症した場合は曝露を想起する（曝露発生の有無にかかわらず）可能性が高い）。 • risk of bias 「高」 の例：曝露データの取得方法が定かでない。 	
<ul style="list-style-type: none"> • 研究の開始時点では懸案のアウトカムが存在していないことに確信が持てない 	
<ul style="list-style-type: none"> • 研究では、懸案のアウトカムに関連する全ての変数に関し、曝露群と非曝露群をマッチさせていない、あるいは統計学的分析により、全ての予後変数について調整を行っていない。 <ul style="list-style-type: none"> • risk of bias 「低」 の例：考えられる予後変数の全てについて、包括的なマッチングまたは調整を行っている。 • risk of bias 「中」 の例：考えられる予後変数の大半について、マッチングまたは調整を行っている。 • risk of bias 「高」 の例：考えられる予後変数のごく一部についてのみ、マッチングまたは調整を行っている、またはマッチングや調整を全く行っていない。両群間に差がない、または両群間の差が統計学的に有意でないことを示す記述が、両群の類似性を確認するためには不十分である。 	
<ul style="list-style-type: none"> • 予後因子の有無の評価に確信が持てない <ul style="list-style-type: none"> • risk of bias 「低」 の例：すべての参加者を対象に面接を実施している。全ての参加者から自記式アンケートへの回答を得ている。カルテをレビューし、再現性を実証している。データベースから予後データを取得しており、予後データ抽出の精度を示す記録がある。 • risk of bias 「中」 の例：カルテをレビューしているが、再現性を実証していない。データベースから予後データを取得しているが、予後データ抽出の質が定かでない。 • risk of bias 「高」 の例：データベースから予後データを取得しているが、予後変数抽出の質に関する記録がない。 	
<ul style="list-style-type: none"> • アウトカムの評価に確信が持てない <ul style="list-style-type: none"> • risk of bias 「低」 の例：盲検化された条件下で独立した研究者が評価を行っている。記録がリンクされている。一部のアウトカム（例：股関節骨折）については、骨折の確認という必要条件を満たすには、カルテの参照で十分である。 • risk of bias 「中」 の例：盲検化されていない条件下で独立した研究者が評価を行っている。自己報告である。一部のアウトカム（例：X線画像の参照が必要な脊椎骨折） 	

<p>については、カルテを参照することではアウトカム評価が十分であるとはいえない。</p> <ul style="list-style-type: none"> • risk of bias 「高」 の例：アウトカム評価が定かでない（説明がない）。 	
<ul style="list-style-type: none"> • コホート群の追跡が不十分である <ul style="list-style-type: none"> • risk of bias 「低」 の例：アウトカムデータの欠損がない。アウトカムデータの欠損の原因が、真のアウトカムに関連している可能性は低い [生存率データの場合、観察打ち切り (censoring) によってバイアスが生じる可能性は低い]。アウトカムデータの欠損の数が各介入群で均衡しており、欠損値の原因が各群で類似している。2 値アウトカムデータの場合、観測されたイベントのリスクと比較し、アウトカムの欠損値の割合が介入効果の推定値に重要な影響を与えるほどではない。連続アウトカムデータの場合、アウトカムの欠損値において考えられる効果サイズ（平均値の差、または標準化平均差）が、観測された効果サイズに重要な影響を及ぼすほど大きくはない。欠損値が適切な方法により補完されている。 • risk of bias 「高」 の例：アウトカムデータの欠損の原因が、真のアウトカムに関連している可能性が高く、各介入群で欠損値の数または原因が一樣ではない。2 値アウトカムデータの場合、観測されたイベントのリスクと比較し、アウトカムの欠損値の割合が、介入効果の推定値に重大なバイアスを生じさせるほど大きいものである。連続アウトカムデータの場合、アウトカムの欠損値において考えられる効果サイズ（平均値の差、または標準化平均差）が、観測された効果サイズに臨床的に意義のあるバイアスを生じさせるほど大きいものである。実際の治療に基づく (as-treated) 解析を、ランダム化の時点で割り付けた介入とは著しく異なる介入を実施した状態で行っている。単純な調整を不適切に適用している可能性がある。 	
<ul style="list-style-type: none"> • 共介入は両群間で類似していない <ul style="list-style-type: none"> • risk of bias 「低」 の例：懸案のアウトカムに影響しうる大半または全ての関連する共介入が、曝露群と非曝露群とで類似していることが記録されている。 • risk of bias 「高」 の例：懸案のアウトカムに影響しうる関連する共介入が、曝露群と非曝露群とで類似していることがほとんど、または全く記録されていない。 	

表4 Risk of bias 以外のグレードダウン要因に関する判定

データの非一貫性 (inconsistency)/ 異質性 (heterogeneity)	はい / 不明 / いいえ *
<ul style="list-style-type: none"> 点推定値が研究間で異なり、その相違がかなり大きい。 各信頼区間の重なりが、ほとんどまたは全くない。 効果の方向が一定ではない。 研究間のばらつきの割合 (I^2) が高い (例, $I^2 > 60\%$ = 高い)。 統計学的な異質性検定で有意 (例, $p < 0.05$)。 	
エビデンスの非直接性 (indirectness)	はい / 不明 / いいえ
<ul style="list-style-type: none"> 対象集団や患者が、われわれが関心のあるものと違う (集団間の差異)。 介入が、われわれが関心のあるものと違う (介入の差異)。 アウトカムが、われわれが関心のあるものと違う (アウトカム指標の差異)。 間接的な比較による結論 (間接比較)。 	
データの不精確さ (imprecision)	はい / 不明 / いいえ
<ul style="list-style-type: none"> 95% 信頼区間に、「効果なし」と、「相当の利益」または「相当の害」が含まれている。 総イベント数やサンプルサイズが少ない (総イベント数 < 300 ~ 400, 総サンプルサイズ < 3000 ~ 4000)。 	
出版バイアス (publication bias)	はい / 不明 / いいえ
<ul style="list-style-type: none"> 小規模研究によってより大きな効果が示されている。 ファンネルプロットの視覚評価あるいは統計学的解析により非対称性がある。 組み込まれた多くの研究が企業からの資金提供を受けている。 発表された試験と未発表の試験との間に結果の乖離がある。 経時的に効果サイズが変化している。 	

*はいが多いとグレードダウンにつながりやすい

表5 グレードアップの3要因に関する判定^d

効果の程度が大きい (large magnitude of effect)	はい / 不明 / いいえ **
<ul style="list-style-type: none"> RR > 2 または < 0.5 (1段階アップ) RR > 5 または < 0.2 (2段階アップ) 	
用量反応勾配 (dose-response gradient)	はい / 不明 / いいえ
<ul style="list-style-type: none"> 用量反応勾配がある場合は、エビデンスの質を1段階上げる。 	
交絡因子の影響 (plausible confounding)	はい / 不明 / いいえ
<ul style="list-style-type: none"> 提示された効果を減弱させている 効果が観察されていないのに当該効果を増大させる方向に働いている 交絡因子の影響がある場合は、エビデンスの質を1段階上げる。 	

**はいが多いとグレードアップにつながりやすい

^d エビデンスの質を低下させる5つの領域のいずれにおいても深刻な限界がみられない場合にのみ、エビデンスの質のグレードアップを判断すべきである。原則的には、グレードアップ要因は観察研究に適用される。

表6 標準様式のエビデンスプロファイルの例

Question: Should Low molecular weight heparin (LMWH) vs. Vitamin K antagonists (VKA) be used for long-term treatment of venous thromboembolism (VTE)?

No. of studies	Quality assessment					No. of patients		Effect		Quality	Importance	
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Clopidogrel	aspirin	Relative (95% CI)			Absolute (95% CI)
Vascular mortality (critical outcome) Fatal MI, fatal ischemic stroke, fatal hemorrhagic stroke and other vascular death (follow up: mean 1.9 years)												
1	randomised trial	serious ¹	not serious	not serious	serious ²	none	372/9599 (3.9%) ³	405/9586 (4.2%) ³	RR 0.92 (0.8 ~ 1.05)	-15/1000 (9 ~ -38) ⁴	⊕⊕⊕ MODERATE	CRITICAL
Nonfatal myocardial infarction (critical outcome) (follow up: mean 1.9 years)												
1	randomised trial	serious	not serious	not serious	serious ²	none	255/9586 (2.7%)	301/9586 (3.1%)	RR 0.85 (0.72 to 1)	-12/1000 (-0 ~ -22) ⁴	⊕⊕⊕ MODERATE	CRITICAL
Nonfatal stroke (critical outcome) Ischemic and hemorrhagic stroke. (follow up: mean 1.9 years)												
1	randomised trial	serious	not serious	not serious	serious ²	none	486/9599 (5.1%) ⁵	528/9586 (5.5%) ⁵	RR 0.92 (0.82 to 1.03)	-9/1000 (3 ~ -20) ⁴	⊕⊕⊕ MODERATE	CRITICAL
Major extracranial bleed (important outcome) Unclearly defined as "any bleeding disorder, severe" (follow up: mean 1.9 years)												
1	randomised trial	serious	not serious	not serious	serious ²	none	132/9599 (1.4%) ⁶	149/9546 (1.6%) ⁶	RR 0.88 (0.7 to 1.12)	-5/1000 (5 ~ -12) ⁷	⊕⊕⊕ MODERATE	IMPORTANT

MD – mean difference, RR – relative risk

- Subgroup analysis suggested no benefit in patients with MI. The subgroup analysis of the composite endpoint reported relative risk reduction of 7.3% for stroke patients, 23.8% for peripheral artery disease patients and relative risk increase of 3.7% for MI patients (test for interaction $p = 0.043$). Based on criteria for credibility we did not believe results from subgroup analysis and therefore we did not rate down for inconsistency.
- Confidence intervals for absolute effects are wide and include important benefit or no benefit/ possible harm with clopidogrel treatment.
- Of the deaths in CAPRIE, 27/405 (6.7%) with aspirin were fatal bleeds and 23/372 (6.2%) with clopidogrel were fatal bleeds.
- Control event rates come from observed events in a meta-analysis of 16 RCTs in secondary prevention (Baigent 2009), adjusted to 5-year time frame.

5. Of the strokes in CAPRIE, 24/486 (4.9%) with aspirin were hemorrhagic and 14/528 (2.6%) with clopidogrel were hemorrhagic.
6. Of the major extracranial bleeds in CAPRIE, 68/149 (45.6%) with aspirin were gastrointestinal and 47/132 (35.6%) with clopidogrel were gastrointestinal.
7. Control event rates come from the observed events in the CAPRIE trial, adjusted to 5-year time frame.
出版社より許可を得て、Vandvikら¹²⁶より転載

表7 ACCP 様式のエビデンスプロファイル例

No. of participants (studies) Follow-up		Quality assessment							No. of patients		
		Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%) With Vitamin K antagonists (VKA) With Low molecular weight heparin (LMWH)	Relative effect (95% CI)	Risk with Vitamin K antagonists (VKA)	Anticipated absolute effects Risk difference with Low molecular weight heparin (LMWH)
Question: Should Low molecular weight heparin (LMWH) rather than the Vitamin K antagonists (VKA) for long-term treatment of Venous thromboembolism (VTE) ? * Bibliography: Low molecular weight heparin compared with vitamin K antagonists for the long-term treatment of venous thromboembolism: a systematic review. Clive Keraron (unpublished) * *											
Overall mortality (follow up: median 6 months)											
2,496 (7 RCTs) 6 months	not serious ¹	not serious	not serious	serious ²	none	⊕⊕⊕ MODER- ATE	202/1231 (16.4%)	204/1265 (16.1%)	RR 0.96 (0.81 ~ 1.13)	Study population 164/1000	-7/1000 (-21 ~ -31)
Recurrent VTE (follow up: median 6 months)											
2,727 (8 RCTs) 6 months	serious ³	not serious	not serious	not serious	none	⊕⊕⊕ MODER- ATE	105/1349 (7.8%) ⁴	67/1378 (4.9%)	RR 0.62 (0.46 ~ 0.84)	no cancer 30/1000	-11/1000 (-5 ~ -16)
Non-metastatic cancer											
80/1000											
-30/1000 (-13 ~ -43)											
Metastatic cancer											
200/1000											
-76/1000 (-32 ~ -108)											
Major bleeding (follow up: median 6 months)											
2,737 (8 RCTs) 6 months	serious ⁹	not serious	not serious	serious ²	none	⊕⊕⊕ LOW	53/1351 (3.9%) ⁵	45/1386 (3.2%)	RR 0.81 (0.55 ~ 1.2)	no cancer 20/1000	-4/1000 (4 ~ -9)

Question: Should Low molecular weight heparin (LMWH) rather than the Vitamin K antagonists (VKA) for long-term treatment of Venous thromboembolism (VTE) ? * Bibliography: Low molecular weight heparin compared with vitamin K antagonists for the long-term treatment of venous thromboembolism: a systematic review. Clive Keraron (unpublished) * *

		Quality assessment					No. of patients		
No. of participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)	Relative effect (95% CI)	Anticipated absolute effects
Follow-up							With Vitamin K antagonists (VKA)	With Low molecular weight heparin (LMWH)	Risk with Vitamin K antagonists (VKA)
									Risk difference with Low molecular weight heparin (LMWH)
									Metastatic cancer
									0/1000
									-0/1000 (-0 ~ -0)
Post thrombotic syndrome (PTS) : Self-reported leg symptoms and signs (follow up: median 3 months)									
1,056 (1 RCT)	serious ⁶	not serious	serious ⁷	not serious	none	⊕⊕○○	200/1000 (20.0%) ⁸	RR 0.85 (0.77 ~ 0.94)	Study population
3 months						LOW	34/56 (60.7%)		200/1000 ⁸
									-30/1000 (-12 ~ -46)

* Limited to LMWH regimens that used 50% or more of the acute treatment dose during the extended phase of treatment.

** Meta-analysis is based on RCTs as referenced in the chapter text. The control event rate for mortality comes from this meta-analysis.

1. One study did not report deaths and could reflect selective reporting of outcomes: Borderline decision.
2. Confidence interval includes important benefit and important harm.
3. None of the studies were blinded.
4. Control event rates from cohort study by Prandoni 2002, adjusted to 6 month time frame.
5. Control event rates from cohort studies by Pradoni 2002 and Beth 1995, adjusted to 6 month time frame.
6. Patients and investigators not blinded.
7. Predictive value at 3 months to long-term is uncertain.
8. Control event rate comes from observational studies in review by Kahn 2004, adjusted to 2 year time frame. All patients wore pressure stockings.
9. None of studies were blinded, but outcomes less subjective: Borderline decision.

表8 SoF の例

慢性閉塞性肺疾患の患者に対する自己管理						
患者あるいは集団：慢性閉塞性肺疾患の患者						
セッティング：プライマリケア, 地域, 外来患者						
介入：自己管理 ¹						
比較：通常ケア						
アウトカム	典型的な比較リスク* (95% CI)		相対効果 (95% CI)	参加者数 (研究数)	エビデンスの質 (GRADE)	コメント
	想定リスク 通常ケア	対応リスク 自己管理				
QOL St George's Respiratory Questionnaire. (SGRQ) スケール (0 ~ 100) 観察期間： 3 ~ 12 か月	対照群の平均 QOL: 38 点 ~ 60 点	介入群の平均 QOL: 2.58 点低 値 (5.14 ~ 0.02 低)		698 (7)	⊕⊕⊕⊕ 中 ²	低スコアは QOL が良いこ とを示す。4 点未満の変化 は患者にとっ ては重要では ないだろう。
呼吸困難 Borg スケール (0 ~ 10) 観察期間： 3 ~ 6 か月	対照群の呼吸困 難： 1.2 ~ 4.1 点	介入群の呼吸困 難：0.53 点低値 (0.96 ~ 0.1 低)		144 (2)	⊕⊕○○ 低 ^{3, 4}	低スコアは改 善を示す。
憎悪の回数と 重症度 ⁵	コメント参照	コメント参照	推定不可 ⁵	591 (3)	コメント参 照	効果は不明。
呼吸関連の入院回 数観察期間： 3 ~ 12 か月	低リスク集団 ⁶		OR 0.64 (0.47 ~ 0.89)	966 (8)	⊕⊕⊕⊕ 中 ⁷	
	10/100	7/100 (5 ~ 9)				
	高リスク集団 ⁶					
	50/100	39/100 (32 ~ 47)				
肺疾患による救急 科への来院回数 観察期間： 6 ~ 12 か月	対照群における救 急科への来院平 均回数：0.2 ~ 0.7 回 / 人 / 年	介入群における救 急科への来院平 均回数：0.1 回多 (0.03 ~ -0.2)		328 (4)	⊕⊕⊕⊕ 中 ⁴	
往診回数 観察期間： 6 ~ 12 か月	対照群における往 診： 1 ~ 5 回 / 人 / 年	介入群における往 診： 0.02 回多い (+1 ~ -1)		629 (8)	⊕⊕⊕⊕ 中 ⁸	

*想定リスクの根拠は脚注で提供される。対応リスク (および 95%CI) は対照群における想定リスク, 介入の相対リスク (および 95%CI) に基づいている。CI: 信頼区間, RR: リスク比

GRADE: GRADE ワーキンググループによるエビデンスの質のグレード

- 「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。
 - 「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、推定値が変わる可能性がある。
 - 「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、推定値が変わる可能性が高い。
 - 「非常に低」 = あらゆる効果推定値が不確実である。
-

1. 自己管理とは、当該疾患向けの治療レジメンの実行に必要なスキルを伝授し、健康に関わる行動の変化を指導し、患者が疾患をコントロールして機能的な生活を送れるよう精神的サポートを提供することを目的とした正式な患者教育プログラムに適用される用語である。14 件の研究のうち、教育の実施形態がグループ教育だったのが 4 件、個別教育だったのが 9 件、書面の教育資料のみだったのが 1 件であった。6 件の研究では、増悪に対する自己治療のための行動計画の使用が評価されていた。
2. 他の 7 研究は統合されず、いくつかの研究においては有意な効果はなかった。
3. 割り付けの隠蔽が 1 研究でなし。不完全なフォローアップ。
4. まばらなデータ
5. 憎悪の定義が異なるため、研究は統合できなかった。
6. 低リスクと高リスクの値は、2 件の研究のコントロール群における入院率の最小と最大の値である（8% は 10%、51% は 50% とした）。
7. 非常に重症の COPD 患者による 2 件の研究がメタアナリシスにおいて重みづけが大きい。
8. 説明のつかない異質性

表9. Evidence to Decision テーブルの例

疑問 / 推奨：最近増悪をきたしたCOPD患者に対しては、通常のコミュニティケアよりも呼吸リハビリテーションを実施すべきか。

集団：最近増悪をきたしたCOPD患者

介入：呼吸リハビリテーション対リハビリテーションなし

セッティング（関連性があれば）：外来患者

決断の領域	判断	判断理由	判断を左右する下位領域
<p>望ましいアウトカムと望ましくないアウトカムのバランス</p> <p>典型的な価値観と好みについての最良推定値を考慮した場合、利益が害と負担を上回る事、または害と負担が利益を上回ることには確信が持てるか。</p>	<p>はい <input type="checkbox"/></p> <p>いいえ <input checked="" type="checkbox"/></p>	<p>相当の望ましい帰結（入院日数の大幅減少、小さいながらも重要な死亡数減少、最小重要差を上回る生活の質の改善を含む）がみとめられ、またこれらの帰結には高い価値が置かれている。望ましくない帰結、不便さ、負担は比較的少なく、最小限の不効用を伴うものにはすぎない。</p>	<p>望ましいアウトカムと望ましくないアウトカムのベースラインリスク：</p> <ul style="list-style-type: none"> サブグループ間でベースラインリスクは類似しているか。 サブグループごとに個別の推奨を提示すべきか。 <p>利益と害の相対リスク：</p> <ul style="list-style-type: none"> 相対的な利益は大きいか。 相対的な害は大きいか。 <p>モデリングの要件：</p> <ul style="list-style-type: none"> これらのアウトカムは多くの推測とモデリングを要するか。 <p>典型的な価値観：</p> <ul style="list-style-type: none"> 典型的価値観にはどのようなものがあるか。 一連の重大なアウトカムの相対的価値観に差はあるか。
<p>効果推定値の確信性（エビデンスの質）</p> <p>「高」または「中」の質のエビデンスはあるか。</p>	<p>はい <input checked="" type="checkbox"/></p> <p>いいえ <input type="checkbox"/></p>	<p>⊕⊕⊕⊙</p> <p>望ましい帰結に関する「中」（死亡、機能、生活の質というアウトカム）～「高」（入院件数）の質のエビデンスが存在し、望ましくないアウトカム（負担）に関する質の高いエビデンスが存在する。</p>	<p>利益とデメリットに関する推定値の確信性、ならびに必要な資源量の推定値の確信性。一部の重大なアウトカムが測定されていない可能性も含め、全ての重大なアウトカムについて検討する。</p> <p>エビデンスのグレードダウンとグレードアップの主な理由。</p>

<p>価値観と好み</p> <p>典型的な価値観と好みについて確信があり、なおかつこれらは標的集団の中で類似しているか</p>	<p>はい <input checked="" type="checkbox"/></p> <p>いい <input type="checkbox"/></p>	<p>患者が入院と死亡を回避して生活の質を改善することに高い価値を置き、リハビリテーションに伴う不便さを回避することに低い価値を置くことに確信が持てる。</p> <p>慢性呼吸器疾患患者の間でこれらの価値観にほとんどばらつきがないことに確信が持てる。</p>	<p>典型的な価値観に関する情報源（一般市民や患者に関するパネルディスカッションや研究）</p> <p>ばらつきなびにばらつき程度の程度に関する推定値の情報源。</p> <p>この推奨に適した、価値観の判断のための方法。</p>
<p>必要資源量への影響</p> <p>必要資源量は、推奨に従うことによって予期される正味の利益に見合ったものか。</p>	<p>はい <input checked="" type="checkbox"/></p> <p>いい <input type="checkbox"/></p>	<p>呼吸リハビリテーションを提供するための必要資源は発生するものの、これらは入院件数の減少による資源ニーズの減少によって相殺され、正味のコストは望ましい帰結に十分見合うものである。</p>	<p>資源単位あたり、どのようなコストがあるか。</p> <p>実行可能性：</p> <ul style="list-style-type: none"> この介入は広く入手可能か。 <p>機会コスト：</p> <ul style="list-style-type: none"> この介入およびその効果には、他の介入から資源を割り当てる価値、または他の介入への資源の割り当てを行わない価値があるか。 <p>セッティング間の差異：</p> <ul style="list-style-type: none"> セッティング間で、資源要件に大きなばらつきがあるか。
<p>推奨の全体的強さ</p>	<p>強い</p>	<p>ガイドラインパネルは、最近増悪をきたした COPD 患者に対する呼吸リハビリテーションを推奨する（注：これは本稿のために提示された架空の推奨であり、診療上の意思決定のために提示されたものではない）。</p>	
<p>エビデンスの統合による推奨の提示</p>	<p>高い価値が置かれるアウトカムへの中等度～大きい効果があることについて中～高の確信性があり、望ましくないアウトカムが軽微であることに中～高の確信性があり、なおかつ望ましくないアウトカムを回避することに高い価値が置かれていないことから、強い推奨を提示できるものと考えられる。</p>		

COPD: chronic obstructive pulmonary disease
 出版社より許可をえて、Andrews ら⁶²より翻訳転載。